THE
PHYSICAL REVIEW
A Journal of Experimental and Theoretical Physics Established by E. L. Nichols in 1893
VOL. 54, NO. 11 DECEMBER 1, 1938 SECOND SERIES

The Angular Spread of Hard Cosmic-Ray Showers

G. WENTZEL
University of Zürich, Zürich, Switzerland
(Received September 27, 1938)

In order to account for the experimental results of Schmeiser and Bothe, the "dynaton" theory has to be modified by the introduction of a "form-factor" which reduces the probability of showers with large angular divergence. A particular invariant "cutting off" rule is discussed.

SCHMEISER AND BOTHE[1] have shown that in the hard cosmic-ray showers, the primaries and secondaries of which are both of the penetrating type, the directions of the shower particles are strongly correlated; the angular divergence of two particles is less than 10° on the average.

From the comparatively large number of these hard showers[2] it may be concluded that the majority of all penetrating cosmic-ray particles are involved in their production. For if only a few percent of the incoming particles, for instance only those of highest energies, were responsible for the hard showers, the cross section of those particles would have to be assumed so large that they could not pass for "penetrating" particles at all; at any rate the observations of Schmeiser and Bothe concerning the connection of the hard showers with the second maximum of the Rossi curve would be difficult to explain in this case. Thus there can hardly be any doubt that the primary radiation which produces the hard showers comprises the bulk of the penetrating cosmic radiation, which, according to present day knowledge, consists of heavy electrons having an average energy between 10^9 and 10^{10} ev. Obviously the secondary particles must be heavy electrons too, since they are distinctly more penetrating than ordinary electrons. It may be admitted, therefore, that the processes in question are nuclear processes in which one heavy electron, with an energy of the order 10^9 or 10^{10} ev in most cases, is absorbed or scattered and several heavy electrons of somewhat lower energy are created ("multiple emission process").

This is, indeed, also in general agreement with theoretical expectation if the heavy electrons are identified with the "dynatons" which, according to Yukawa's theory, give rise to the nuclear forces. The cross sections of the processes in question, as estimated by Heitler[3] on the basis of Yukawa's theory, are so large at energies of the order 10^9 ev that dynatons of about this energy may well be assumed to be the primaries of the hard showers.

In this case, however, as Heitler[3] remarked already, there arises a difficulty as to the explanation of the angular correlation of the hard shower particles. For, according to the unmodified quantum theory, a pronounced intensity maximum in

[1] Schmeiser and Bothe, Ann. d. Physik **32**, 161 (1938).
[2] Professor Bothe kindly informed the writer that the number of hard showers produced in 17 cm of lead (second maximum of the Rossi curve) is rather like, and certainly not much inferior to, the number of the ordinary soft (cascade) showers produced in 1.7 cm of lead (first maximum).
[3] Heitler, Proc. Roy. Soc. **166**, 529 (1938).

the forward direction is to be expected, just as in the case of Compton scattering, only if the momentum p of the incident particle is large in comparison with Mc (M=mass of the scattering proton-neutron), and the spread angle will then be of the order $(Mc/p)^{\frac{1}{2}}$. If this angle were to be as small as 5 or 10° as found by Schmeiser and Bothe, the primary energy would have to be assumed of the order 100 Mc^2 or 10^{11} ev, in contradiction with the foregoing estimation; at any rate it is hard to believe that the energy spectrum of the penetrating radiation has sufficient intensity above 10^{11} ev[4] to account for the many hard showers observed. (For this reason, Heitler[3] considers the possibility that the hard showers originate in impacts with electrons rather than with nuclei; but the assumptoin of so large an interaction between heavy and ordinary electrons can hardly be reconciled with the long lifetime of the β-radiators and of the heavy electron itself.)

Yet the observed angular correlation cannot be said to contradict the dynaton interpretation of the hard showers absolutely, since it is well known from other applications that Yukawa's theory, like quantum electrodynamics, must be assumed to break down in problems involving very large energies. This is usually expressed by saying that the high energy states have to be "cut off." As to the question where the cut should be made, the only information available up to now appears to be that resulting from theoretical speculations on the masses and the magnetic moments of the proton and the neutron, which are rather doubtful; *a priori*, of course, such a limitation remains arbitrary to some extent. Now the observations of Schmeiser and Bothe may be considered to furnish direct experimental evidence in this respect: The nonoccurrence of hard showers with larger angular divergence may have the simple meaning that these showers are forbidden by certain "cutting off" rules.

This interpretation may be supported by the following reasoning. As is well known, the high energy states are canceled automatically if the elementary particles are assumed to have a nonvanishing spatial extension, and the canceling rules are closely connected with the assumptions about the structure and the dimensions of the particles. In the mathematical formalism they manifest themselves in "form factors" which are, e.g., in the case of scattering, functions of $(r/\lambda)\sin\vartheta/2$, where r denotes the linear dimension of the scattering system, $\lambda=\hbar/p$ the wavelength and ϑ the scattering angle. The form factor equals 1 if $(r/\lambda)\sin\vartheta/2\ll 1$; this defines the field of validity of the ordinary theory ($r=0$). On the other hand the processes with $(r/\lambda)\sin\vartheta/2\gg 1$ are almost excluded. At high energies ($\lambda\ll r$) the domain of the allowed scattering angles is $\lesssim\hbar/rp$. If one identifies r with the classical electron radius (e^2/mc^2) or with the Compton wavelength of the proton $(\hbar/Mc)$, the average scattering angle will be about 5° if the energy is of the order 10^9 or 10^{10} ev, respectively.

Of course the question remains open whether or how far there is a physical meaning in speaking of the spatial extension of elementary particles. The tentative introduction of form factors F may therefore only have the significance that the equation $F=1$ defines the domain in which the unmodified quantum theory holds true.

In the present case particularly, it is readily seen that a too straightforward introduction of form factors into the Hamiltonian does not help to remove the above-mentioned difficulty. Let the matrix element of the multiple emission process be expressed in the usual manner by products of the matrix elements of the corresponding single absorption and emission processes. If, then, the form factors representing the spatial structure of the proton and the neutron are introduced into the matrix elements of the *single* processes, the only success will be that the energy of each absorbed or emitted particle will be limited, but there will be no restriction as to the angles between the directions of the various particles.

However, the desired angular correlation may be obtained by multiplying the total matrix element of the compound process, irrespectively of the virtual intermediate states, with a "form factor" saying that the amount of the total change of momentum of the heavy particle (proton-neutron) shall be limited:

$$|\Delta\mathbf{P}|\lesssim\hbar/r, \tag{1}$$

i.e., that the large changes of momentum

$$|\Delta\mathbf{P}|\gg\hbar/r$$

[4] Compare Euler, Naturwiss. 26, 382 (1938).

are forbidden. For, if one applies this to high energy collisions, the total momentum of the emitted particles will equal the momentum of the incident particle within an accuracy of the order $\hbar/r$, and consequently the expectation value of the angle between two emitted particles will be of the order $\hbar/rp$, in accordance with observation.

In order to be Lorentz invariant $|\Delta\mathbf{P}|$ should be measured in an invariant reference system, such as the system in which the total momentum vanishes, or the system in which the energy of the proton-neutron remains constant ($\Delta E=0$). These two systems do not, in general, coincide, but provided that $\hbar/r \lesssim Mc$, the limits defined by (1) will be practically the same whether $|\Delta\mathbf{P}|$ is measured in the one or in the other reference system or also, say, in the initial rest system of the proton-neutron. If (1) holds in the system where $\Delta E=0$, this condition reads

$$|\Delta\mathbf{P}|^2-(\Delta E/c)^2 \lesssim (\hbar/r)^2 \qquad (2)$$

in an arbitrary reference system.

A rule which establishes a discrimination between the various particles involved may appear strange at first sight. But it must be remembered that, according to the theory in question, the proton-neutron actually plays a privileged role, in that it remains present during the whole process, through all intermediate virtual phases. Thereupon one should also expect that the restrictive rule must be essentially the same when the particles involved are replaced by particles of similar nature (e.g., the proton-neutron by an electron, or a heavy electron by a photon). Thus, for instance, a shower of heavy electrons produced by an incident *photon* of corresponding energy should have the same angular spread, assuming that the "cutting-off" radius r is independent of the nature of the particles to be absorbed or emitted. Actually Schmeiser and Bothe[1] find that the hard showers appearing at the *first* maximum of the Rossi curve, most of which are probably created by the soft (photon) component of the cosmic radiation, have about the same angular distribution as those connected with the second maximum.

Another point to be discussed is the bearing of the form factor on the total cross section of the heavy electron. As has been observed by several authors, the probability of the multiple emission processes in question must be expected to be relatively large, compared with a multiple photon emission, for two reasons: Firstly[5] the parameter α, the powers of which define the successive approximations in the perturbation method, is probably rather large ($\alpha\sim 1/10$), at any rate considerably larger than the corresponding parameter in quantum-electrodynamics ($e^2/\hbar c \cong 1/137$); this can be deduced from the absolute magnitude of the nuclear forces. Secondly,[6] the matrix element of the interaction of the dynaton with the proton-neutron, provided that it is chosen such that the proton-neutron force gets the right spin character (essentially Majorana force), contains terms with an energy dependence which is by one degree higher than in the interaction of the photon with the electron in quantum-electrodynamics. Because of this the theoretical cross sections of high multiple emission processes increase very rapidly with increasing energy as long as no form factors are introduced, rather like the cross sections of multiple β-emission processes as discussed by Heisenberg on the basis of Fermi's β-theory.

For the following computation the heavy electron may be assumed to have the spin 1, for this is the most probable value according to the theories of the proton-neutron force.[6, 7] If such a particle, endowed with an energy $E \gg \mu c^2$ (μ = mass of the heavy electron), collides with a proton or neutron, the theoretical cross section of the process in which n heavy electrons are emitted into the solid angles $d\Omega_1, d\Omega_2 \cdots d\Omega_n$ ($1 \leqq n \ll E/\mu c^2$) has the order of magnitude

$$\alpha^{n+1}\left(\frac{\hbar}{\mu c}\right)^2\left(\frac{E}{\mu c^2}\right)^{2n} d\Omega_1 d\Omega_2 \cdots d\Omega_n, \qquad (3)$$

at least for small scattering angles ϑ. Without form factors, this holds for any energy as far as $\vartheta \sim (Mc^2/E)^{\frac{1}{2}}$; the integral cross section would therefore be of the order

$$\alpha\left(\frac{\hbar}{\mu c}\right)^2\left(\alpha\frac{M}{\mu}\frac{E}{\mu c^2}\right)^n,$$

[5] Wentzel, Naturwiss. **26**, 273 (1938).
[6] Kemmer, Proc. Roy. Soc. **166**, 127 (1938); Bhabha, Proc. Roy. Soc. **166**, 501 (1938).
[7] Yukawa, Sakata, Taketani, Proc. Phys.-Math. Soc. Japan **20**, April, 1938; Fröhlich, Heitler, Kemmer, Proc. Roy Soc. **166**, 154 (1938); Stueckelberg, Helv. Phys. Acta **11**, 245 (1938).

and heavy electrons with energies above 10^9 ev would no longer be "penetrating," which, of course, can hardly be true. If, however, a form factor is added which allows only scattering angles $\vartheta \lesssim \hbar c/rE$, the integral cross section turns out to be of the order

$$\alpha\left(\frac{\hbar}{\mu c}\right)^2\left(\alpha^{\frac{1}{2}}\frac{\hbar}{\mu c}\frac{1}{r}\right)^{2n}. \tag{4}$$

In this case the cross section will not only become constant with increasing energy but also the dependence on n will be comparatively weak, so that showers with many particles ($1 \ll n < E/\mu c^2$) can be expected to occur with an accordingly reduced probability. The order of magnitude of the total cross section of all processes in question will not be very much larger than the cross section for simple scattering, which is obtained by putting $n=1$ in (4); also this value seems to be in substantial agreement with the experimental data.

Thus, the "canceling rule" as suggested above can be said to be well adapted to the observed facts, including the angular correlation as well as the absolute frequency of the hard showers.

Recently Heisenberg has expressed the opinion that the quantum-theoretical formulae for the transition probabilities should be applicable only to such processes in which the invariant changes of momentum $|\,|\Delta\mathbf{P}|^2-(\Delta E/c)^2|^{\frac{1}{2}}$ remain below the finite limit $\hbar/r$ for *each* particle involved in the process.[8] If $\hbar/r < \mu c$, the absorption and emission of heavy electrons would then be altogether beyond the confines of the present theory. According to the above results, however, it seems rather, that the range of validity of the theory is *larger* than Heisenberg assumes, in that the restrictive condition applies *only* to the change of momentum of the *proton-neutron*, or, speaking more generally, to that particle which remains present during the whole compound process.

If several such particles cooperate in a process, the restricting rule must naturally be applied to each of them.[9] So, for instance, in the case of the proton-neutron collision, large changes of momentum of each particle will be forbidden, and this means, obviously, that the potential of the proton-neutron force remains finite at zero distance, the wave-lengths $\lambda \ll r$ being canceled in the Fourier integral. This warrants the finiteness of all quantities related to the proton-neutron interaction, such as collision cross sections or the binding energy of the deuteron, even in higher approximations of the perturbation method.

On the other hand, self-energies and similar quantities (e.g., the magnetic moments which can be calculated from the self-energy of the proton or neutron in an external magnetic field) are, obviously, not affected by the form factor proposed, and thus remain infinite. (But in this respect it would be of no use either to extend the restrictive condition (2) to the *single* virtual processes, since for instance the left-hand expression in (2) vanishes identically in the case of an absorbed or emitted light quantum and therefore the electromagnetic self-energy would remain unaltered.) Presumably infinities of this type can only be eliminated by means of a subtraction formalism such as has proved indispensable in the "hole theory" of the positron. It is generally admitted that this problem is closely connected with other fundamental problems the solution of which will require an improved knowledge of the phenomena occurring at high energies. The hope may be justified that such knowledge will be supplied by further cosmic-ray experiments and their theoretical examination.

In conclusion, the writer wishes to express his admiration and his gratefulness to Professor Arnold Sommerfeld on the occasion of his seventieth birthday.

[8] Similar ideas seem to dominate in a new theory suggested by Wataghin (compare the preliminary notes in Nature **142**, 393 (1938) and Comptes rendus **207**, 358 (1938)). A consistent explanation of all cosmic-ray phenomena will scarcely be possible on these lines.

[9] If the electron-positron-pair creation is interpreted according to the "hole theory," the question arises whether the particle which is lifted from a negative to a positive energy state is perhaps also subject to a rule like (2) (with the reversed sign of the left-hand expression).

Remarks on Nuclear Disintegrations by Cosmic Rays

W. HEITLER
H. H. Wills Physical Laboratory, University of Bristol, England
(Received September 22, 1938)

I

IN the Wilson chamber nuclear disintegrations by cosmic rays have occasionally been observed by Anderson and Neddermeyer[1] and by Brode and Starr.[2] A few heavy particles (probably protons) emerge from a single point in all directions, in many cases accompanied by a few electron or heavy electron tracks. The process is, however, so rare that not more than about ten cases have been observed so far. A more suitable way of studying those nuclear disintegrations is the direct photographic method which has been used recently for this purpose by Blau and Wambacher.[3] On a photographic plate α-ray tracks or α-ray *and* proton tracks (according to the type of plates and development) are recorded automatically whereas light tracks (electrons and heavy electrons) escape observation. A nuclear disintegration appears on the plate as a little star.

Blau and Wambacher have started to investigate in this way the nuclear disintegrations systematically. Most of the protons emitted in a disintegration have energies of 1–10 Mev. The number of particles ranges between 1 and 10. Occasionally also higher energies have been observed. Furthermore it seems that the yield increases very much with the height above sea level, only a very small number of disintegrations have been observed at sea level so far.

How can we now picture such a nuclear disintegration? The problem can be divided well into two different questions:

The first question is: Which is the process and its probability by which a cosmic-ray particle, e.g., an electron, light quantum or heavy electron, can transfer a large amount of energy to the nucleus? In general, the energy will be transferred in the first instant of time only to one of the nuclear particles, e.g., a proton or a neutron.

The second question is then: How is the energy of the nuclear particle distributed among the other nuclear particles, and in which way can the excited nucleus fall down again to its ground level; which particles are then emitted?

II

We start with a discussion of the second question. In general the nuclear particle, which has received a large amount of energy from the cosmic particle, will at once interact with the other nuclear particles and transfer a part of its energy to them. It has, however, been pointed out by Heisenberg[4] that there are also cases in which the nuclear particle can leave the nucleus undisturbed. This is the case if the particle in question lies on the surface of the nucleus and receives a velocity directed outwards. If, on the other hand, the velocity is directed inwards it will interact with other nuclear particles. This gives rise to a certain heating up of the nucleus.

The discussion by Bohr and Kalckar[5] of collision processes of neutrons with nuclei in the energy region of 1–10 Mev has shown that in this case the *whole* primary energy is transformed into nuclear heat energy and thermal equilibrium is reached before any of the nuclear particles can leave the nucleus. It is not *a priori* certain whether this is also the case for cosmic-ray energies (100–1000 Mev). In the first instant of time only a certain portion of the nucleus (the part which the original particle has hit) obtains a very high temperature. The probability for releasing a nuclear particle increases very rapidly with the temperature. It depends therefore on the "heat conductivity" of the nucleus whether thermal equilibrium is reached or whether some of the nuclear particles are released

[1] Anderson and Neddermeyer, Phys. Rev. **50**, 263 (1936).
[2] Brode and Starr, Phys. Rev. **53**, 3 (1938).
[3] Blau and Wambacher, Akad. Wiss. Wien **146**, 623 (1937).
[4] Heisenberg, Sächsische Akad. Wiss. 369 (1937).
[5] Bohr and Kalckar, Ugl. Danske, Vid. Selsk **14**, 10 (1937).

from the very hot part of the nucleus *before* the equilibrium is restored. The heat conductivity can hardly be calculated exactly, but a simple consideration shows that the time necessary to release a nuclear particle can at least not be shorter than the relaxation time for restoring the thermal equilibrium. The latter will be of the order of magnitude of the time between two collisions of two nuclear particles. The time necessary for a particle to leave the nucleus reaches, for a very hot nucleus, a minimum value which is also equal to the time necessary to traverse the nucleus. This minimum value is only reached for temperatures larger than the whole binding energy of the nucleus (or of the very hot part of the nucleus). We see, therefore, that any nuclear particle will have to suffer at least some collisions before it can leave the nucleus and it seems probable that at least in many cases, complete thermal equilibrium is restored before any particles can leave the nucleus.

Let us consider more in detail the case of complete equilibrium. This case has been treated satisfactorily by Weisskopf.[6] If the total excitation energy is still smaller than the total binding energy all his assumptions are also valid for cosmic-ray energies. For a heavy nucleus the binding energy is of the order of 2000 Mev. We can therefore apply his results for cosmic-ray energies of 100–1000 Mev.

According to Weisskopf the temperature of the nucleus is given by

$$T=(aE)^{\frac{1}{2}}, \quad a \text{ between } 0.05 \text{ and } 0.2 \text{ Mev}. \quad (1)$$

For E of the order 100–500 Mev we obtain a temperature of 2–10 Mev. As long as the temperature is so high it is to be expected that most of the particles leave the nucleus with energies of the order T. After having released some particles the nucleus cools down and then also slower particles are to be expected. This result is in good agreement with the experiments by Blau and Wambacher. There are, however, also cases in which particles with higher energies than 10 Mev have been observed. It might be that this is due to the release of particles from a *partially* heated nucleus, as discussed above.

[6] Weisskopf, Phys. Rev. **52**, 295 (1937).

The total number of particles to be expected would be given by E/T if the temperature would remain constant. E/T is therefore only a minimum number. For $E=100$ Mev we obtain about 20 particles. At least half of them (probably more) are, however, neutrons. Thus we expect on the average about 5–10 protons. This figure is also in reasonable agreement with Blau and Wambacher's experiments.

III

A far more difficult, and at present hardly answerable, question is that of the energy transfer of a cosmic-ray particle to the nucleus. It is not even known which of the two main constituents of cosmic radiation, the soft or the hard component, give the main effect.[7] It seems that the rate of nuclear disintegrations increases rapidly with height above sea level. This result, if established, would indicate that the effect is mainly due to the soft component. The question can only be settled by measuring the absorption of the radiation producing the disintegrations.[8]

Let us consider the question from the theoretical point of view. A nuclear excitation can only be caused by an interaction of the cosmic-ray particle with a nuclear particle. The long range Coulomb forces can hardly produce an appreciable effect. Heisenberg,[4] and more recently Williams,[9] have treated the effect assuming short range forces of the type of the nuclear forces. However, those forces exist only if the cosmic-ray particle itself is a proton or neutron. Although heavy particles occur (very rarely) in cosmic rays, we do not think that the effect can be ascribed entirely to them for the following reason. There is so far no indication for the existence of *primary* protons in cosmic radiation. The heavy particles found there are probably secondaries produced by electrons or heavy electrons, and, as secondaries, they can only be produced in some sort of nuclear disintegration. The question reduces therefore to the

[7] Possibly one will have to take into account also other components of cosmic radiation such as neutrons, etc., which, for the general behavior of cosmic rays, play a less important role.

[8] Those absorption measurements are in progress by Blau (according to a kind private communication) and by Heitler and Powell.

[9] Williams, Nature **142**, 341 (1938).

question in which way a nuclear disintegration can be produced by an electron (light quantum) or heavy electron, or, in other words, whether there is a short range interaction between the latter particles and a nuclear particle.

The question is indeed one of the fundamental problems of modern quantum theory. There is a direct interaction between an electron and a nuclear particle together with a neutrino which is responsible for the β-decay, according to Fermi's theory or Yukawa's version of this theory. From the β-decay we have, however, knowledge about this interaction only for energies of the order of 1 Mev. If we could extrapolate this interaction for higher energies—we discuss this question below—it would still be far too weak to account for the observed rate of nuclear disintegrations by cosmic rays.

A very strong interaction exists between a heavy electron and a nuclear particle. According to the modern theories of the nuclear forces and the magnetic moments of the heavy particles[10, 11] processes of the type

$$\begin{aligned} N+Y^{+}&\rightarrow P \\ P+Y^{-}&\rightarrow N \end{aligned} \qquad (2)$$

(Y=heavy electron, P=proton, N=neutron) can occur in analogy to the emission and absorption of a light quantum by an electron. The "charge" g of the heavy particles responsible for the processes (2) is so large that $g^2/\hbar c \backsimeq \frac{1}{6}$ in contrast to $e^2/\hbar c = 1/137$.

The application of this theory to collisions of heavy electrons with nuclei leads to all sorts of processes—absorption and scattering of heavy electrons, etc. All these processes will have comparatively large cross sections. Moreover, if a heavy electron is scattered by a (free) nuclear particle, it is possible that multiple processes can occur as soon as the energy is larger than a few hundred Mev.[12] Thus, instead of one scattered heavy electron a number of heavy electrons can be emitted if a primary heavy electron hits the nucleus (multiple shower). In all these processes a certain amount of energy is transferred to the nuclear particle which is used for a nuclear excitation in the way described above.[13] Of course, only in a small fraction of cases the amount of energy transferred to the nucleus is sufficiently large to give rise to a nuclear disintegration. For the cross section for the various scattering processes a rough estimate has been given by Heitler.[12] For energies of the order of 100 Mev it is of a reasonable order of magnitude.

It can thus be understood that *heavy electrons* give rise to nuclear disintegrations. The light particles which have frequently been observed to accompany nuclear disintegrations may well be the heavy electrons emitted in a multiple process as described above. Ordinary electrons and light quanta could also interact with nuclear particles by means of an intermediate emission of heavy electrons. The cross section should be $e^2/\hbar c$ times smaller than the corresponding one for a heavy electron. The soft cosmic-ray component should therefore not give rise to an appreciable number of disintegrations. It would therefore be difficult to understand the rapid increase of the rate of nuclear disintegrations with height on grounds of this theory.

IV

The theory of the heavy electron and its interaction with nuclear particles meets with very fundamental difficulties for energies high compared with the rest energy of the heavy electron (*viz.* 50–100 Mev). The very existence of multiple processes shows that the ordinary methods of quantum theory can no longer be applied in this energy region. The interaction between a heavy electron and a nuclear particle is of the form[10, 11]

$$(g/\lambda)\ \mathrm{div}\ \psi(0)+(f/\lambda)\sigma\ \mathrm{curl}\ \psi(0), \qquad (3)$$

where g and f are two constants of the order of 5 electron charges, $\lambda=mc/\hbar$ (m=mass of the heavy electron), $\psi(0)$ the wave function of the heavy electron at the position of the nuclear

[10] Yukawa, Sahata and Taketani, Proc. Math. Phys. Soc. Japan **20**, 319(1938).
[11] Fröhlich, Heitler and Kemmer, Proc. Roy. Soc. **166**, 154 (1938). Cf. also: Bhabha, Proc. Roy. Soc. **166**, 501 (1938), Stückelberg, Phys. Rev. **52**, 41 (1937).
[12] Heitler, Proc. Roy. Soc. **166**, 529 (1938).

[13] A direct absorption of a heavy electron by a nucleus can only take place if the nuclear particle is bound. The cross section decreases very rapidly with the ratio binding energy : primary energy (analogous to the photoelectric effect). According to calculations by Arley (for which I am very much indebted to Dr. Arley), the probability for absorption is negligible for cosmic-ray energies. Only scattering processes will therefore give rise to a nuclear excitation.

particle and σ the spin of the nuclear particle. The interaction (3) can successfully explain the properties of the proton and neutron. However, for a free heavy electron with wave number k (3) can be considered as the first term in an expansion of a more general interaction with respect to k/λ, e.g., for *small* energies of the heavy electron. There is no reason whatsoever to expect that the same form of the interaction (3) also holds for higher energies $k>\lambda$. In fact, if we apply (3) for higher energies the theory leads to divergences which are even more serious than the well-known divergences occurring in quantum electrodynamics. In all probability, the right interaction between a heavy particle and a heavy electron will be of a *nonlinear type* and this fact marks the limits of applicability of the present quantum mechanics. The same considerations apply to the β-interaction between electrons and nuclear particles.

Thus, the problem of the interaction between heavy electrons (and electrons or light quanta) and nuclear particles for high energies seems to be a very fundamental one, leading beyond the limits of the present quantum theory. Fortunately, it seems that this problem is well accessible to experiments. The behavior of heavy electrons in traversing matter, in particular the nuclear disintegrations caused by them, provide a direct experimental test for the interaction in question.

This was not so in quantum electrodynamics. There the departure from the linear laws has only found an expression in quite general facts such as the finite rest mass of the electron and the merely theoretical difficulties occurring in higher order calculations. All the radiation effects of an electron could very well be treated by a first-order approximation theory and there are no experiments giving any indication for higher order effects.

It seems that experiments on the behavior of the hard cosmic-ray component could, for the first time in physics, open an insight in the laws of physics which are beyond the validity of the quantum theory.

DECEMBER 1, 1938 PHYSICAL REVIEW VOLUME 54

The Zeeman Effect in the Spectrum of Argon

J. B. GREEN AND B. FRIED
Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio

(Received October 3, 1938)

New measurements on the Zeeman effect of about 140 lines of the spectrum of argon have been completed, yielding g values for several complete configurations, in particular the $3p^5 6s$, $3p^5 7s$, $3p^5 4d$, and $3p^5 5d$, and several levels of the $3p^5 6d$. These have been compared with the g values calculated by means of the quantum mechanics. In some cases the agreement is good, but g sums show discrepancies. These are discussed in detail. The measurements differ in several cases from those of other observers. Concrete evidence of the jj parentage of some of the levels is presented.

THE classification of the argon spectrum was practically completed by Meissner.[1, 2] His classifications were slightly modified in a few instances by Rasmussen,[3] who also extended the classifications to include some weak lines not given by Meissner.

Work on the Zeeman effect was begun by Bakker[4] who reported the g values for the $3p^5 4s$ configurations, and carried on by Pogany,[5] who investigated the $3p^5 4p$ configurations, and Terrien and Dykstra,[6] who investigated the $3p^5 4p$ and $3p^5 5p$ configurations. Since then, Jacquinot[7] has reported some measurements on the $3p^5 ms$

[1] Meissner, Ann. d. Physik **51**, 115 (1916).
[2] Meissner, Zeits. f. Physik **39**, 172 (1926); **40**, 839 (1927).
[3] Rasmussen, Zeits. f. Physik **75**, 695 (1932).
[4] Bakker, Nature **126**, 955 (1930).
[5] Pogany, Zeits. f. Physik **93**, 364 (1935).
[6] Terrien and Dykstra, J. de Phys. **5**, 439 (1934).
[7] Jacquinot, Comptes rendus **206**, 1635 (1938).

TABLE I. *Summary of the measurements of the $3p^54p$ configuration.*

λ	COMB.	J_1J_2	PATTERN	g_1	g_2
*9224.50	$1s_2-2p_6$	1, 2	(0), (0.202), −, −, *1.508*	1.104	1.306
*9122.98	$1s_5-2p_{10}$	2, 1	(0), (0.475), *1.029*, 1.505, 1.994	1.507	1.987
8667.94	$1s_3-2p_7$	0, 1	(0), 0.836	0/0	0.836
8521.442	$1s_2-2p_4$	1, 1	(0.283), 0.820, 1.109	1.107	0.821
8424.648	$1s_4-2p_8$	1, 2	(0), (0.285), *0.821*, 1.119, 1.404	1.401	1.115
8408.213	$1s_2-2p_3$	1, 2	(0), (0.170), −, 1.271, *1.429*	1.101	1.265
8265.524	$1s_2-2p_2$	1, 1	(0.288), 1.102, 1.385	1.101	1.387
†8115.309	$1s_5-2p_9$	2, 3	(0), (0.169), (0.368), *0.979*, 1.187, 1.368	1.518	1.338
8103.692	$1s_4-2p_9$	1, 1	(0.569), 0.839, 1.402	1.403	0.838
8014.785	$1s_5-2p_8$	2, 2	(0.383), [0.752], *1.110*, *1.489*, 1.878	1.489	1.110
7948.176	$1s_3-2p_9$	0, 1	(0), 0.823	0/0	0.823
7724.210	$1s_3-2p_2$	0, 1	(0), 1.390	0/0	1.390
7723.759	$1s_5-2p_7$	2, 1	(0), (−), −, 1.507, *2.131*	1.507	0.833
7635.107	$1s_5-2p_6$	2, 2	(−), (−), 1.091, *1.308*, *1.506*, 1.726	1.513	1.302
7514.650	$1s_4-2p_5$	1, 0	(0), 1.399	1.399	0/0
7503.868	$1s_2-2p_1$	1, 0	(0), 1.099	1.099	0/0
7471.18	$1s_4-2p_4$	1, 1	(0.590), 0.822, 1.411	1.410	0.822
7383.978	$1s_5-2p_1$	2, 1	(0), (−), −, 1.508, *2.200*	1.508	0.816
7147.042	$1s_4-2p_3$	1, 2	(0), (0.152), *1.109*, 1.259, −	1.410	1.260
7067.218	$1s_5-2p_3$	2, 2	(0.213), (*0.507*), 0.983, *1.250*, *1.508*, 1.779	1.504	1.264
6965.431	$1s_5-2p_2$	2, 1	(0), 1.622	*1.501*	1.378
6677.282	$1s_4-2p_1$	1, 0	(0), 1.402	1.402	0/0

* Not previously reported. † Not reliable.

configurations, with which the present work is in disagreement, and Lörinczi[8] has made some calculations of the g factors to be expected in the $3p^54d$ and $3p^55d$ configurations, together with measurements of a few argon lines.

The present investigation covers the region λλ5000–9700 of the spectrum of argon, and was carried out with the same apparatus and under the same conditions as reported in a previous paper,[9] on the spectrum of neon, with this slight difference which we think is worthy of note. In the case of the discharge in neon, a pressure of about 7 mm was found to operate best, while in the case of argon, in order to produce a strong discharge, a pressure of 0.7 mm was found satisfactory. This is just the opposite set of conditions that usually exists in these two gases outside a magnetic field. Even under these circumstances, the spectrum of argon II is very strongly excited and changing the pressure, up or down, did very little to alter the relative intensity of arc and spark spectra. While the part of the discharge tube between the poles was intensely blue, the rest of the tube glowed red, and this gave a very simple means of separating arc from spark spectrum.

Exposures were usually about 48 hr.; Eastman special spectroscopic plates were used and developed in Edwal 12 developer.

Table I is a summary of the measurements of the $3p^54s-3p^54p$ multiplet, and is given for the sake of completeness. Pogany[5] measured a few of these lines in the parallel polarization, assuming Bakker's[4] results for the g factors of the $3p^54s$ configurations, while Terrien and Dykstra[6] were not able to resolve some of them. These measurements yield the results given in Table II. Wave-lengths and classifications are by Meissner. The agreement is within the limit of error. The results for the $3p^54p$ configuration will be summarized in a later table.

Table III contains a summary of the measurements of lines not previously reported. Except in the few cases where Zeeman effect measurements indicate a preference for Rasmussen's[3] classifications (indicated by an asterisk), the

TABLE II. *Summary of results for $3p^54s$ configuration.*

	AUTHORS	BAKKER	CALCULATED
$1s_2$	1.102	1.100	1.101
$1s_4$	1.404	1.400	1.399
Σg	2.506	2.500	2.500
$1s_5$	1.506	1.50	1.500

[8] Lörinczi, Zeits. f. Physik **109**, 175 (1938).
[9] Green and Peoples, Phys. Rev. **54**, 602 (1938).

TABLE III. *Summary of the measurements of lines of argon not previously reported.*

λ	COMB.	$j_1 j_2$	PATTERN	g_1	g_2
8053.33	$2p_6-4d_5$	2, 1	(*0*), (0.173), *1.132*, −, −	1.305	1.478
8046.08	$2p_7-4d_6$	1, 0	(0), 0.838	0.838	0/0
7891.10	$2p_6-4d_3$	2, 2	(−), (*0.250*), −, −, −, −	*1.305*	1.430
7484.24	$2p_7-4d_1''$	1, 2	(0), 0.943	*0.838*	0.908
7435.33	$2p_6-3s_5$	2, 2	(−), (*0.385*), −, −, −, −	*1.305*	1.498
7425.24	$2p_3-4s_1'''$	2, 3	(0), 0.938	*1.260*	1.099
7412.31	$2p_4-4s_1''$	1, 2	(0), 1.152	*0.818*	0.985
7392.97	$2p_6-3s_4$	2, 1	(0), 1.431	*1.305*	1.179
7372.12	$2p_9-4d_4'$	3, 4	(0), 1.021	*1.333*	1.255
7353.32	$2p_8-4d_4$	2, 3	(0), 1.040	*1.112*	1.076
7350.78	$2p_2-3s_2$	1, 0	(0), 1.377	1.377	0/0
7316.00	$2p_2-3s_2$	1, 1	(0.121), 1.330	1.390	1.270
7311.71	$2p_7-3s_4$	1, 1	(0.344), 0.834, 1.179	0.834	1.179
7206.99	$2p_3-3s_2$	2, 1	(0), 1.255	*1.260*	1.270
7162.57	$2p_5-4s_1'$	0, 1	(0), 0.876	0/0	0.876
7158.83	$2p_4-3s_3$	1, 0	(0), 0.817	0.817	0/0
7125.80	$2p_4-3s_2$	1, 1	(0.443), 0.817, 1.274	0.821	1.271
7107.50	$2p_8-3s_5$	2, 2	(0.402), [0.757], −, *1.105*, *1.488*, 1.860	1.105	1.488
7086.70	$2p_5-3s_2$	0, 1	(0), 1.274	0/0	1.274
7030.25	$2p_9-3s_5$	3, 2	(*0*), (0.167), (0.338), *1.015*, 1.194, −, −, −	1.347	1.512
6951.46	$2p_6-4s_1''$	2, 2	(−), (0.653), −, 0.994, −, 1.644	1.319	0.993
6937.67	$2p_{10}-4d_6$	1, 0	(0), 1.984	1.984	0/0
6888.17	$2p_7-4s_1''''$	1, 2	(*0*), (0.213), −, −, *1.271*	0.844	1.057
6879.59	$2p_7-4s_1''$	1, 2	(0), 1.080	*0.838*	0.999
6871.29	$2p_{10}-4d_5$	1, 1	(0.517), 1.461, 1.979	1.979	1.461
6766.56	$2p_6-4s_1'$	2, 1	(*0*), (0.424), −, 1.307, *1.726*	1.305	0.883
6752.83	$2p_{10}-4d_3$	1, 2	(*0*), (0.540), *0.894*, 1.438, 1.985	1.980	1.437
6719.20	$2p_5-5d_5$	0, 1	(0), 1.395	0/0	1.395
6698.85	$2p_6-3s_2$	2, 1	(0), 1.328	*1.305*	1.259
6664.02	$2p_8-4s_1''$	2, 2	(−), (*0.244*), 1.046	1.107	0.984
6660.64	$2p_7-3s_3$	1, 0	(0), 0.847	0.847	0/0
6632.04	$2p_7-3s_2$	1, 1	(0.429), 0.849, 1.270	0.847	1.272
6604.85	$2p_8-4s_1'''$	2, 3	(0), 1.165	*1.112*	1.139
6596.10	$2p_9-4s_1''$	3, 2	(*0*), (0.359), (0.711), −, −, 1.334, 1.704, *2.061*	1.343	0.984
6594.66	$2p_1-4s_2$	0, 1	(0), 1.286	0/0	1.286
6538.12	$2p_9-4s_1'''$	3, 3	(0.187), (0.395), (0.611), −, 0.934, *1.129*, *1.332*, 1.537, 1.758	1.338	1.133
6513.84	$2p_4-4s_4$	1, 1	(0.332), 0.811, 1.171	0.817	1.164
6493.97	$2p_8-4s_1'$	2, 1	(*0*), (0.241), −, 1.120, *1.359*	1.119	0.879
6481.15	$2p_5-4s_4$	0, 1	(0), 1.158	0/0	1.158
6466.56	$2p_5-5d_2$	0, 1	(0), 0.820	0/0	0.820
6431.57	$2p_8-3s_2$	2, 1	(*0*), (−), *0.964*, −, −	*1.112*	1.260
6416.31	$2p_{10}-3s_5$	1, 2	(*0*), (0.482), *1.019*, 1.505, 1.988	1.985	1.502
6384.72	$2p_{10}-3s_4$	1, 1	(0.801), 1.187, 1.991	1.990	1.188
6369.58	$2p_6-5d_5$	2, 1	(0), 1.251	*1.305*	1.413
6364.89	$2p_7-5d_6$	1, 0	(0), 0.833	0.833	0/0
6349.20	$2p_{10}-4d_2$	1, 1	0.768, (1.210), 1.994	1.986	0.768
6309.14	$2p_7-5d_5$	1, 1	(0.555), 0.844, 1.399	0.844	1.399
6307.66	$2p_6-5d_3$	2, 2	(*0.161*), 1.343	1.303	1.383
6296.88	$2p_2-5s_1''$	1, 2	(*0*), (−), *1.157*, −, −	*1.380*	1.269
6248.40	$2p_7-5d_3$	1, 2	(*0*), (0.540), −, 1.382, *1.914*	0.844	1.380
6243.39	$2p_2-6d_6$	1, 0	(0), 1.364	1.364	0/0
6215.95	$2p_3-5s_1''$	2, 2	(0), 1.269	*1.260*	1.278
6212.51	$2p_6-5d_1'$	2, 3	(0), 1.113	*1.305*	1.209
6173.10	$2p_7-5d_1''$	1, 2	(0), 1.054	*0.838*	0.946
6170.18	$2p_6-4s_5$	2, 2	(−), (*0.403*), 1.090, *1.306*, *1.527*, 1.744	1.305	1.512
6165.11	$2p_3-5s_1''''$	2, 2	(0.411), [0.872], 1.260, 1.698	1.260	0.822
6155.23	$2p_6-4s_4$	2, 1	(0), 1.441	*1.305*	1.169
6145.43	$2p_3-5s_1'''$	2, 3	(0), 0.890	*1.260*	1.075
6128.71	$2p_2-6d_3$	1, 2	(0), 1.078	*1.380*	1.179
6127.38	$2p_8-5d_5$	2, 1	(*0*), (0.305), *0.830*, 1.120, −	1.122	1.418
6119.67	$2p_4-6d_5$	1, 1	(0.413), 0.810, 1.230	0.812	1.228
6113.47	$2p_7-4s_5$	1, 2	(*0*), (0.676), −, 1.517, *2.187*	0.842	1.514
6105.64	$2p_4-5s_1''''$	1, 2	(0), 0.799	*0.818*	0.806
6104.60	$2p_4-6d_6$	1, 0	(0), 0.811	0.811	0/0
6101.16	$2p_2-4s_2$	1, 1	(0), 1.337	*1.380*	1.294
6098.81	$2p_7-4s_4$	1, 1	(0.323), 0.826, 1.170	0.831	1.165
6090.76	$2p_5-6d_5$ $2p_1-8d_5$	0, 1 0, 1	(0), 1.232	0/0	1.232
6085.86	$2p_7-5d_2$	1, 1	(0), 0.819	*0.838*	0.800

TABLE III—*Continued.*

λ	COMB.	$j_1 j_2$	PATTERN	g_1	g_2
6081.23	$2p_2-6d_2$	1, 1	(0.212), 1.281	1.387	1.175
6059.38	$2p_{10}-4s_1''''$	1, 2	(0), [0.985], 1.988	1.988	—
6052.73	$2p_{10}-4s_1''$	1, 2	(0), [0.995], 1.989	1.989	—
6043.22	$2p_8-5d_4$	2, 3	(0), 1.038	*1.112*	1.075
6032.13	$2p_9-5d_4'$	3, 4	(0), 1.011	*1.333*	1.253
6025.14	$2p_3-4s_2$	2, 1	(0), 1.231	*1.260*	1.289
6013.68	$2p_9-5d_3$	3, 2	(0), 1.299	*1.333*	1.367
6005.74	$2p_3-6d_2$	2, 1	(0), 1.293	*1.260*	1.194
5999.00	$2p_8-5d_1''$	2, 2	(*0.317*), 1.020	1.099	0.941
5987.29	$2p_9-5d_4$	3, 3	(−), (0.521), [0.805], *1.082*, *1.338*, 1.604, 1.892	1.344	1.076
5981.90	$2p_8-5d_1'$	2, 3	(0), 1.355	*1.112*	1.193
5971.59	$2p_4-4s_3$	1, 0	(0), 0.823	0.823	0/0
5968.31	$2p_4-4s_2$	1, 1	(0.474), 0.821, 1.298	0.822	1.297
†5964.46	$2p_5-5s_1'$	0, 1	(0), 0.850	0/0	0.850
5949.26	$2p_4-6d_2$	1, 1	(0.375), 0.816, 1.192	0.816	1.192
5942.67	$2p_8-4s_5$	2, 2	(0.413), (*0.754*), −, *1.115*, *1.498*, 1.876	1.120	1.498
5940.86	$2p_5-4s_2$	0, 1	(0), 1.295	0/0	1.295
5928.82	$2p_8-4s_4$	2, 1	(0), 1.090	*1.112*	1.156
5927.13	$2p_9-5d_1'$	3, 3	(*0.303*), 1.248	*1.305*	1,191
5916.58	$2p_8-5d_2$	2, 1	(*0*), (−), −, 1.091, *1.387*	1.091	0.795
5912.09	$2p_{10}-4s_1'$	1, 1	0.873, (1.116), 1.984	1.985	0.871
5882.63	$2p_1-3s_2$	1, 0	(0), 1.999	1.999	0/0
5860.31	$2p_{10}-3s_2$	1, 1	(0.714), 1.266, 1.981	1.981	1.266
5834.27	$2p_6-5s_1''$	2, 2	(0), 1.284	*1.305*	1.263
5802.08	$2p_6-6d_5$	2, 1	(0), 1.350	*1.305*	1.215
5772.12	$2p_6-5s_1'''$	2, 3	(0), 0.785	*1.305*	1.132
5739.52	$2p_7-5s_1''''$	1, 2	(0), 0.788	*0.838*	0.805
5738.40	$2p_7-6d_6$	1, 0	(0), 0.824	0.824	0/0
5712.48	$2p_5-7d_5$	0, 1	(0), 1.321	0/0	1.321
5689.91	$2p_2-6s_1''$	1, 2	(0), 1.215	*1.380*	1.273
*5689.64	$2p_6-6d_3$	2, 2	(*0.191*), 1.244	1.292	1.196
5681.90	$2p_6-6d_1'$	2, 3	(0), 1.184	*1.305*	1.244
5659.13	$2p_6-5s_5$	2, 2	(0.379)	*1.305*	1.495
5650.71	$2p_{10}-5d_6$	1, 0	(0), 1.984	1.984	0/0
5648.66	$2p_6-6d_2$	2, 1	(0), 1.369	*1.305*	1.177
*5641.34	$2p_7-6d_3$	1, 2	(*0*), (0.378), −, 1.216, *1.594*	0.838	1.216
‡5639.11	$2p_7-5s_1'$	1, 1	(0), 0.838	*0.838*	0.838
5635.54	$2p_7-6d_1''$	1, 2	(0), 1.241	*0.838*	1.107
5623.76	$2p_3-6s_1''$	2, 2	(0), 1.260	*1.260*	1.260
5620.89	$2p_7-4s_3$	1, 0	(0), 0.824	0.824	0/0
5617.97	$2p_7-4s_2$	1, 1	(0.480), 0.841, 1.307	0.844	1.304
5606.74	$2p_{10}-5d_5$	1, 1	(0.578), 1.391, 1.985	1.981	1.395
5597.46	$2p_3-6s_1'''$	2, 3	(0), 0.915	*1.260*	1.087
5588.69	$2p_8-5s_1''''$	2, 2	(0.639)	*1.112*	0.793
5581.83	$2p_9-5s_1''$	3, 2	(0), 1.407	*1.333*	1.259
5572.55	$2p_8-5s_1'''$	2, 3	(0), 1.128	*1.112*	1.120
5558.71	$2p_{10}-5d_3$	1, 2	(*0*), (0.592), *0.808*, 1.391, 1.985	1.988	1.398
5523.93	$2p_9-5s_1'''$	3, 3	(0.628)	*1.333*	1.124
5506.11	$2p_8-6d_4$	2, 3	(0), 0.989	*1.112*	1.051
5495.88	$2p_9-6d_4'$	3, 4	(0), 1.027	*1.333*	1.256
5457.37	$2p_8-6d_2$	2, 1	(0), 1.044	*1.112*	1.248
5451.66	$2p_{10}-4s_5$	1, 2	(*0*), (0.481), *1.023*, 1.501, −	1.981	1.501
5443.21	$2p_5-7d_3$	2, 2	(0), 1.274	*1.305*	1.243
5442.22	$2p_9-6d_1'$	3, 3	(0), 1.291	*1.333*	1.248
5439.97	$2p_{10}-4s_4$	1, 1	(0.814), 1.168, 1.982	1.982	1.168
5373.49	$2p_7-7d_1''$	1, 2	(0), 1.035	*0.838*	0.970
5254.48	$2p_7-6s_1''''$	1, 2	(0), 0.746	*0.838*	0.777
5252.80	$2p_8-7d_4$	2, 3	(0), 1.045	*1.112*	1.078
5187.75	$2p_{10}-5s_1''$	1, 2	(*0*), [0.616], −, 1.265, *1.975*	1.975	1.265
5162.29	$2p_{10}-6d_6$	1, 1	(0.742), 1.239, 1.985	1.984	1.239
5151.40	$2p_{10}-6d_5$	1, 0	(0), 1.981	1.981	0/0
5118.20	$2p_8-6s_1'''$	2, 3	(0), 1.105	*1.112*	1.109
5087.09	$2p_8-8d_4$	2, 3	(0), 0.954	*1.112*	1.032
5054.18	$2p_{10}-4s_2$	1, 1	(0.687), 1.298, 1.977	1.978	1.296
5048.81	$2p_{10}-5s_5$	1, 2	(*0*), (0.491), *1.016*, 1.493, −	1.982	1.499
4887.95	$2p_{10}-7d_6$	1, 1	(0.624), 1.357, −	1.981	1.357

Note:—Squared bracket [ ] means overlapped parallel and perpendicular components.
* Given by Meissner as $2p-5s_1'$.
† Not classified by Meissner.
‡ Improperly classified by Meissner as $2p_4-6s_4$.

TABLE IV. *Summary of results for the $3p^5 6s$ and $3p^5 7s$ configuration.*

$3p^5 6s$					$3p^5 7s$				
	TERM VALUE	g CALC.*	g OBS. (AUTHORS)	g OBS. (JACQUINOT)		TERM VALUE	g CALC.*	g OBS. (AUTHORS)	g OBS. (JACQUINOT)
$3s_2$	5950.13	1.311	1.271	1.36	$4s_2$	3229.21	1.326	1.296	1.30
$3s_4$	7351.29	1.189	1.184	1.18	$4s_4$	4632.09	1.174	1.164	1.15
Σg		2.500	2.455	2.54	Σg		2.500	2.460	2.45
$3s_5$	7428.39	1.500	1.500	1.50	$4s_5$	4671.41	1.500	1.506	1.50

Note: Jacquinot observed for the $5s_5$ term of the $3p^5 8s$ g value 1.499.

* Houston, Phys. Rev. **33**, 297 (1929).

TABLE V. *Summary of results for the $3p^5 4p$ configuration.*

LEVEL	J	g (CALC.)*	g (POGANY)	g (T. & D.)	g (AUTHORS)[a]	g (AUTHORS)[b]
$2p_2$	1	1.363	1.379	1.37	1.380	1.388
$2p_4$		0.887	0.819	0.815	0.818	0.820
$2p_7$		0.774	0.840	0.825	0.838	0.837
$2p_{10}$		1.976	—	—	1.984	1.987
	Σg	5.000	—	—	5.020	5.032
$2p_3$	2	1.280	1.248	1.26	1.260	1.260
$2p_6$		1.271	1.302	1.30	1.305	1.304
$2p_8$		1.116	1.121	1.11	1.112	1.113
	Σg	3.667	3.671	3.67	3.677	3.677
$2p_9$	3	1.333	—	—	1.338	1.339

[a] Averages from $2p-md$ lines. [b] Averages from $1s-mp$ lines. * Green, Phys. Rev. **52**, 736 (1937).

TABLE VI. *$3p^5 4d$ configuration of argon. Parameters for the least square calculation are: $F_0=-1064.6$; $F_2=86.39$; $G_1=69.80$; $G_3=5.60$; $\zeta_p=893.05$; $\zeta_d=3.85$.*

LEVEL	ENERGY LEVEL (ABSOLUTE)	ENERGY LEVELS (CALC.)[a]	(CALC.)[b]	ENERGY LEVELS (OBS.)	g (CALC.)[b]	g (CALC.)[a]	g (OBS.)[a]
d_6	8599.40	7.56		0.0	0/0	0/0	0/0
d_5	8460.07	124.7	115	131.3	1.29	1.461	1.467
d_2	7263.70	1328.7	1321	1335.7	0.86	0.766	0.768*
s_1'	6099.53	2524.0	2430	2499.9	0.84	0.773	0.877
				$J=1$	Σg 2.99	3.000	3.112
d_3	8204.85	387.9	381	394.6	1.30	1.433	1.437*
d_1''	7666.55	942.4	919	932.9	0.87	0.864	0.908†
s_1''	6492.46	2055.7	2060	2106.9	1.12	0.875	0.987
s_1''''	6510.60	2193.6	2108	2088.9	1.05	1.161	1.057*
				$J=2$	Σg 4.34	4.333	4.389
d_4	7898.59	639.5	644	700.8	1.083	1.075	1.077†
d_1'	[7546]‡	1114.6	935	[1053]‡	1.210	1.190	—
s_1'''	6357.99	2204.7	2276	2241.4	1.110	1.152	1.133
				$J=3$	Σg 3.403	3.417	—
d_4'	8087.81	449.1		511.6		1.250	1.255†

[a] Authors. [b] Lörinczi. * Based on one resolved line. † Based on one unresolved line. ‡ Meissner lists no $4d_1'$.

wave-lengths and classifications are as given by Meissner.[2]

Using only resolved patterns to determine the g factors, we find for the $3p^5 6s$ and $3p^5 7s$ configurations the results given in Table IV.

There can be no question but that Jacquinot's[7] measurement of $3s_2$ is in error. The value indicated by him is based on measurements of only one line, which appears as a weak line just on the verge of resolution. The g values we have indicated are averages based on well-resolved lines, five for $3s_2$ and $4s_2$, three for $3s_4$ and four for $4s_4$, and when used to determine "blend"[10] patterns for unresolved lines yield results in agreement with observed patterns, as may be seen by consulting Table III.

The accuracy of the g values indicated in Table IV is at least one-half percent and we are therefore compelled to consider the discrepancy in the g sums as being significant. Both the $3p^5 6s$ and the $3p^5 7s$ configurations show g sums for the $j=1$ levels which fall far short of the predicted theoretical values. This discrepancy can only be accounted for as a result of interaction between the $3p^5 ms$ and $3p^5 md$ configurations. We shall discuss this in greater detail later on.

Table V is a summary of the results for the $3p^5 4p$ configuration. This has been reported before, but is given here to show the agreement among the several observers, and does include the g value of $2p_{10}$ which has not hitherto been recorded.

The agreement between the authors' two sets of values is in all cases within the one-half percent we usually allow ourselves, and in most cases better than that, but the results in column a are more reliable than b; the latter based, except for $2p_3$, on only two or three lines; and in the case of $2p_{10}$ the value 1.987 comes from only one line. But between observers, in the case of $2p_3$, Pogany's[5] value 1.248 appears too small, and the value 0.825 given by Terrien and Dykstra[6] appears too small. In both cases we found it was necessary to use values for $2p_3$ and $2p_7$ indicated by our own measurements in order to get agreement between observation and calculation of "blend" patterns.

Mr. J. F. Eichelberger of this laboratory has calculated parameters for the $3p^5 4d$ configuration of argon by the method of least squares, and has kindly allowed us to publish the following data for that configuration. Table VI lists the g values calculated from his data.

Table VII is a summary of observed and calculated data for $3p^5 5d$. No parameters have been calculated for $3p^5 6d$, and although a large number of lines involving this configuration appeared on our plates, most of them were not resolved patterns. We have, however, listed the g values obtained from these patterns in Table VIII.

To begin with, the following corrections and additions to Meissner's classifications are suggested by these measurements. The level called $5s_1'^{(j=1)}$ by Meissner[2] has clearly a j value of 2 and together with the other levels already assigned to $3p^5 6d$ by Meissner, supplies the missing $6d_3$ level of this configuration. This assignment is substantiated by the very satisfactory g sum. To supply the $5s_1'$ level, Rasmussen[3] has suggested the level 3296, not classified by Meissner. This suggestion is verified by the measurements which indicate that it has a j value equal to 1 and the g value agrees very well with the calculated g value. We were unable to obtain any lines involving either Meissner's $6s_1'$ or Rasmussen's $6s_1'$ and were therefore unable to make any decision with respect to these levels. Meissner has omitted the level $4d_1'$ from his list of classifications. The calculated position of this level is 7485. This would yield a line in combination with p_6 in the region of λ7475. At 7500.70 is a line questioningly classified as $2p_3 - 4s_1''$, by Meissner. The discrepancy between observed and calculated positions is 0.88 cm^{-1}, which is much too large. We have therefore assumed this line to be the missing $2p_6 - 4d_1'$ and have listed the corresponding value of 7546 for $4d_1'$ in Table VI to compare with the calculated position of this level. Other lines in Meissner's list might have been chosen, but this is the only one which gives the correct relative position of $4d_1'$ with respect to $4d_1''$.

The situation with respect to the levels ms_1'' and ms_1'''' is of very particular interest. Both of these levels have j values of 2 and are separated from each other in jj coupling only as a result of the ζ_d separation; that is, by the

[10] Shenstone and Blair, Phil. Mag. **8**, 765 (1929).

TABLE VII. *Observed and calculated data for the $3p^5 5d$ configuration.*

LEVEL	ENERGY LEVEL (ABSOLUTE)	ENERGY LEVELS (CALC.)[10]	(CALC.)[b]	ENERGY LEVELS (OBS.)	g (CALC.)[b]	g (CALC.)[a]	g (OBS.)[a]
d_6	5317.39	38		0.0	0/0	0/0	0/0
d_5	5178.63	117	124	139	1.457	1.412	1.400
d_2	4597.22	765	830	720	0.734	0.758	0.813
s_1'	3296.04‡	1976	1950	2021	0.815	0.830	0.846
				$J=1$	Σg 3.006	3.000	3.059
d_3	5024.56	265	264	293	1.363	1.365	1.387§
d_1''	4829.28	460	502	488	0.939	0.938	0.941*
s_1''	3738.54	1692	1638	1579	0.895	1.239	1.265*‖
s_1''''	3605.97	1638	1669	1711	1.138	0.791	0.802†
				$J=2$	Σg 4.335	4.333	4.395
d_4	4951.29	339	345	366	1.067	1.086	1.076*
d_1'	4781.81	557	585	536	1.215	1.205	1.199
s_1'''	3554.04	1762	1730	1763	1.125	1.127	1.127†
				$J=3$	Σg 3.407	3.418	3.402
d_4'	5075.38	244	—	242	—	1.250	1.253†

[a] Authors. [b] Lörinczi.
‡ Not classified as $5s_1'$ by Meissner.
* Based on one resolved line.
§ Lörinczi gives 1.40.
† Based on unresolved patterns.
‖ Lörinczi gives 0.929.

influence of the *md*-electron. Yet, in spite of this, the g values of the two levels, which approach each other as m increases, approach the g values for jj coupling, namely 1.267 and 0.767 for the $\frac{1}{2}$, $\frac{5}{2}$ and the $\frac{1}{2}$, $\frac{3}{2}$ levels respectively. Theoretically the $\frac{1}{2}$, $\frac{5}{2}$ level should lie above the $\frac{1}{2}$, $\frac{3}{2}$ level and associated with it should be the larger g value. We find in the case of argon $3p^5 5d$ and $3p^5 6d$ that the larger g value is associated with the lower level, in both cases called s_1'', and therefore consistent with the neon spectrum. In the case of $3p^5 4d$, however, the larger g value is also associated with the lower level but this is called $4s_1''''$ by Meissner. To be consistent with the rest of the assignments in the argon spectrum we should call this level $4s_1''$, and vice versa, even though this inverts the whole system, and shows that some serious perturbation is at work, since ζ_d should be essentially positive, instead of negative, as the positions of these levels would indicate. This is the case in several of the ^2D and ^2F terms of alkali-like spectra. We believe that this is the first time that definitive information regarding the jj coupling genealogy has ever been obtained. Incidentally, the calculations of the g values indicate that Sampson's[11] correlations of $5s_1''$ and $5s_1''''$ are incorrect, and we have corrected them in Table VII.

We are unable to account for the fact that the observed and calculated values given by Lörinczi[8] differ so markedly from our own, especially when his calculated term-values are about as good as ours, and in some cases even closer to the observed values. He has given a set of general formulas for the calculation of g values for $p^5 d$ configurations, and we have not taken the trouble to check it, because such a formula, we have found, serves very little purpose. It is much easier to find the g values from the original matrix, making each case a special one.

The spectrum of argon, like that of all the rare gases, has each configuration divided into two groups, an upper, converging toward $^2P_{\frac{1}{2}}$ and a lower, converging toward $^2P_{3/2}$. In addition to this, the *md* and $(m+2)s$ configurations occupy practically the same positions. Furthermore, in particular, the upper group of the $3p^5 5d$ and the lower group of the $3p^5 6d$ and the upper group of $3p^5 7s$ and the lower group of $3p^5 8s$ all lie between 3175 cm^{-1} and 3738 cm^{-1}. We should therefore expect the g sums to be very seriously perturbed since these configurations have the same parity. We have already remarked a negative discrepancy of 0.045 in

[11] Sampson, Phys. Rev. **52**, 1157 (1937).

TABLE VIII. *Summary of results for the $3p^5 6d$ configuration.*

LEVEL	J	g (OBS.)	LEVEL	J	g (OBS.)
$6d_2$	1	1.186	$6d_1'$	3	1.245
$6d_5$	1	1.233*†	$6d_4$	3	1.052
$6s_1'$	1	—	$6s_1'''$	3	1.098
$6d_1''$	2	1.107‡		Σg	3.395
$6d_3$	2	1.206*‡			
$6s_1''$	2	1.264	$6d_4'$	4	1.256
$6s_1''''$	2	0.777			
	Σg	4.354			

* Based on resolved patterns.
‡ Meissner calls this level $5s_1'$.
† Lörinczi finds 1.49 for $6d_5$.

the $J=1$ g sum of $3p^5 6s$ and of 0.040 in the $j=1$ g sum of $3p^5 7s$. In the $j=1$ g sum of the $3p^5 4d$ configuration, we find a positive discrepancy of 0.112 mainly attributable to $4s_1'$ if we compare observed and calculated values. In the $j=1$ g sum of $3p^5 5d$ we find a positive discrepancy of 0.059 mainly attributable to $5d_2$. If these discrepancies are real and not the result of experimental error, they certainly tend to compensate each other, as they should. A complete check, however, would necessitate a complete set of data on the g values of the whole spectrum, including the $3p^5 3d$ and $3p\ 5s$ configurations, and this is not available at present, and there seems to be little hope for it in the near future, at least with the present apparatus. We are inclined to believe that these are real discrepancies, although their accuracy may be as poor as ± fifty percent, if we consider that even for resolved patterns, an estimated ± one-half percent experimental error is reasonable. In the case of $j=2$ g sums, we find positive discrepancies of 0.056 for the $3p^5 4d$, mainly attributable to $4d_1''$ (the g sum of $4s_1''$ and $4s_1''''$ is about the same for calculated and observed, although neither g value by itself is a good fit, owing undoubtedly to the poor agreement of observed and calculated energy level positions), and 0.062 for the $3p^5 5d$. There is no discrepancy in the only $j=2$ levels of $3p^5 6s$ and $3p^5 7s$ that cannot be accounted for by experimental error. The situation here is somewhat different from the case of $j=1$, for here the g values in a number of cases were determined from unresolved patterns, yielding a maximum accuracy of ± one percent. There is a chance, therefore, that these discrepancies may be due to experimental error. The $j=3$ g sums are certainly within experimental error, and if the *ms* and *nd* configurations were the sole ones to influence each other, we should expect no perturbation in this g sum for there are no levels in the *ms* configurations with $j=3$.

In conclusion, we wish to state that it is indeed a great pleasure to dedicate this work to Professor Arnold Sommerfeld, the former teacher of one of us during his student days at the University of Wisconsin, and his ever-present inspiration.

Note by J. F. Eichelberger, of Ohio State University, added in proof at the request of the authors:—The following table gives the results of a least-squares solution of the energy levels of the $3p^5 6d$ configuration of argon, together with the g values calculated therefrom. The attempt to find a set of parameters which give a satisfactory set of solutions for these energy levels has not been very successful; in fact, the first approximation is almost as good as the second. This lack of success must be attributed to interaction between the $3p^5 6d$ configuration and $3p^5 5d$ and $3p^5 7s$ and $3p^5 8s$. The evidence for this suggestion is seen in the fact that the d_4 and d_6 levels are inverted with respect to their usual theoretical positions, and the calculated g value for d_3 shows poorest agreement. Further, the agreement of the g values for $j=3$ are much better than any of the others, and it is here that we should expect the smallest perturbation since there are no levels with $j=3$ in the $3p^5 ms$ configurations.

Calculated and observed g values and energy levels for the $3p^5 6d$ configuration of argon.

LEVEL	ENERGY LEVELS (ABS.)	(CALC.)	(OBS.)	g (CALC.)	g (OBS.)
d_6	3602.55	−14.7	0.0	0/0	0/0
d_4	3643.51	+45.3	−41.0	1.412	1.233
d_2	3175.53	436.5	427.0	0.797	1.186
S_1'	1825.33	1743.8	1777.2	0.791	——
				Σg 3.000	——
d_3	3302.90	169.8	299.7	1.359	1.206
d_1''	3284.65	330.4	317.9	0.946	1.107
S_1''	2045.03	1612.3	1557.5	1.244	1.264
S_1''''	1998.09	1582.3	1604.5	0.784	0.777
				Σg 4.333	4.354
d_4	3337.54	223.0	265.0	1.079	1.052
d_1'	3279.03	385.1	323.5	1.218	1.245
S_1'''	1961.50	1660.3	1641.0	1.120	1.098
				Σg 3.417	3.395
d_4'	3458.28	142.7	144.3	1.250	1.256

$F_0 = -657.78$; $F_2 = +29.43$; $G_1 = +23.32$; $G_3 = +2.274$;
$\zeta_p = -920.66$; $\zeta_d = +4.11$.

Resonance Broadening of Spectral Lines

W. V. Houston
California Institute of Technology, Pasadena, California
(Received October 7, 1938)

A general treatment of the collision of an excited atom with an unexcited one shows that the three principal methods of calculating collision broadening are all justified as approximations to the complete treatment. When the fine structure of the lines is taken into account it follows that all members of a Russell-Saunders multiplet are broadened to the same extent. The calculated broadening turns out to be $\Delta\nu=(\frac{8}{3})^{\frac{1}{2}}NS(J, J')/\hbar(2J+1)(2J'+1)$, which is less by a factor of five than that observed in the work of Hughes and Lloyd.

THE broadening of spectral lines which accompanies an increase in the pressure of the absorbing or emitting gas has been the subject of a large number of theoretical investigations from various points of view.[1] Usually the principal attention has been given to the broadening produced by a foreign gas, and most of the reliable experimental work has been of this kind. The broadening in a homogeneous gas, known as resonance broadening, is very much greater, but its study has been accompanied with such technical difficulties that until recently very few dependable measurements have been made on it. The recent work of Hughes and Lloyd,[2] however, provides experimental data of some reliability, and it is therefore of interest to examine again the theoretical predictions as to the absolute magnitude of the effect, and as to the relative broadening of the components of a multiplet.

The various theoretical points of view from which this subject has been treated show considerable diversity. Some of them have been definitely shown to be wrong, but there remain several others which have sometimes been thought to be distinct, even though they lead to approximately the same results. Furthermore, in all of the work which has been published there appears to be no examination of the relative broadening of the different members of a multiplet, although it might be surmised on general grounds that all such lines would be broadened to the same extent. It is proposed to show here, first, that the various methods of approach which have been used on this problem are merely different approximations to the complete treatment; and, second, that the broadening can be expressed in such a way as to show that the different members of a Russell-Saunders multiplet are broadened to the same extent.

The theoretical points of view from which this problem has been treated may be classified under three headings as follows:

1. Furssow and Wlassow[3] have emphasized the fact that when an excited atom passes by an unexcited atom of the same kind there is a finite probability that the energy of excitation will be transferred from the first to the second atom, without the intervention of radiation. This results in a reduction of the lifetime of the excited atom and a corresponding broadening of the line radiation emitted. Although this statement is in terms of an emission line, the conclusion applies to an absorption line as well.

2. A second method of treatment is closely related to the Lorentz theory of collision broadening in which the phase of a classical oscillator is regarded as being changed in a random way during a collision. From this point of view the problem is to compute the amount of the change of phase in terms of the forces of interaction and the closeness of approach. The Fourier analysis of this interrupted oscillation gives the familiar dispersion curve. In the quantum mechanical treatment the Fourier analysis is made of exp $(i/\hbar)(E'-E)t$, where E' is the energy of the excited atom and E is the energy of the ground state. The quantity E' is considered to vary

[1] For summaries of the theoretical work see: Margenau and Watson, Rev. Mod. Phys. **8**, 22 (1936); V. Weisskopf, Physik. Zeits. **34**, 1 (1933).

[2] Hughes and Lloyd, Phys. Rev. **52**, 1215 (1937).

[3] Furssow and Wlassow, Physik. Zeits. der Sowjetunion **10**, 379 (1936). The criticism of Weisskopf's work contained in this paper does not appear to be justified.

during the collision, by an amount which can be computed, and so to produce a change in phaes. This method has been extensively used by Weisskopf and by Lenz.[4]

3. A third method of treatment is to consider the motion of the atomic nuclei under the attractive or repulsive force due to the electronic interaction between the two atoms. If the stationary states of this motion are found, and if it is assumed that the electronic motion is essentially unchanged during the collision, it is possible to compute the probability of transition, from the excited to the ground state, with the emission of radiation and a change in the kinetic energy of the atoms. This method shows much more clearly than the others that the conservation of energy is made possible through the change in the kinetic energy of the atomic motion; and Weisskopf has shown that this treatment leads, at least approximately, to the same result as the method of Fourier analysis. This method has recently been treated extensively by Jablonski.[5]

Since all three of these methods lead to roughly the same result, it is easy to suppose that they are really equivalent. On the other hand, and especially since the result seems to be considerably smaller than required by experiment, it might be supposed that these are three effects which, in the first approximation, are independent and must therefore be added together. A complete analysis, however, shows that they are merely three approximations to the exact quantum mechanical treatment.

General Theory of a Collision

The theory as usually treated, and as treated here, applies to the case of low pressures only. Pressures are low when the mean free time between collisions is considerably greater than the duration of a collision. Under these circumstances it is sufficient to treat simple collisions between two atoms only, and to add together the effects of the different collisions. This has as a consequence that the broadening will be proportional to the number of molecules per unit volume.

Consider then two identical atoms which for simplicity will be treated as having only two electronic states. They will be capable of occupying a continuum of states of translational motion, but apart from the translational states there will be three important electronic states.

$$\begin{aligned} U_a &= u_0(1)u_1(2) \\ U_b &= u_0(2)u_1(1) \\ U_0 &= u_0(1)u_0(2). \end{aligned} \tag{1}$$

The state u_0 is the ground state and the state u_1 the excited state. The arguments of these functions are the electronic coordinates of the first and second atoms respectively. The states U_a and U_b differ with respect to the number of the atom which is excited, but they have the same energy.

To treat the problem of the emission of radiation, the state of the radiation field must also be taken into account. It may be assumed that when the atomic state is U_a or U_b there is no radiation in the field, and when the atomic state is U_0 there is one quantum of energy $h\nu$. Let the function describing the field in the first case be F_0 and in the second case be F_ν. Then the three states of importance are

$$\begin{aligned} \psi_a &= U_a v_1(R, t) F_0 \exp\left[-(i/\hbar)E_1 t\right] \\ \psi_b &= U_b v_1(R, t) F_0 \exp\left[-(i/\hbar)E_1 t\right] \\ \psi_0 &= U_0 v_0(R, t) F_\nu \exp\left[-(i/\hbar)(E_0+h\nu)t\right]. \end{aligned} \tag{2}$$

The function $v(R, t)$ is a function of the distance between the atomic nuclei and the time; $v_1(R, t)$ and $v_0(R, t)$ represent the nuclear motion before and after the collision, respectively. For certain methods of treatment these functions will represent wave packets and for others they will be functions corresponding to a precise value of the translational energy. They could, of course, be written as functions of the coordinates of the two atoms separately, but a function of the relative coordinate R gives all of the generality which is of importance.

The Hamiltonian function for this system will contain terms which represent the atomic energies, the translational energy, and the energy of the radiation field. In addition there will be a term $V(R)$, which represents the dipole interaction energy between the two atoms and is a function of their separation R, and another term S which represents the interaction between the electrons and the radiation field.

[4] W. Lenz, Zeits. f. Physik **80**, 422 (1932).
[5] A. Jablonski, Acta Physica Polonica **6**, 371 (1937).

Let the general solution of the problem be

$$\Psi = a\psi_a + b\psi_b + c\psi_0, \tag{3}$$

where a, b, and c are functions of the time. Substitution of this form into the Schrödinger equation leads to

$$\begin{aligned}
\dot{a} &= -(i/\hbar)(a|V|b)b \\
&\quad -(i/\hbar)(a|S|c)c \exp\left[(i/\hbar)(E_1-E_0-h\nu)t\right] \\
\dot{b} &= -(i/\hbar)(b|V|a)a \\
&\quad -(i/\hbar)(b|S|c)c \exp\left[(i/\hbar)(E_1-E_0-h\nu)t\right] \\
\dot{c} &= -(i/\hbar)(c|S|a)a \exp\left[-(i/\hbar)(E_1-E_0-h\nu)t\right] \\
&\quad -(i/\hbar)(c|S|b)b \exp\left[-(i/\hbar)(E_1-E_0-h\nu)t\right].
\end{aligned} \tag{4}$$

For the cases of interest the matrix elements of V are of more importance than those of S. This corresponds to the fact that the broadening due to pressure is greater than the natural width of the lines. Hence, to find an approximate solution, neglect at first the matrix elements of S altogether. The remaining two equations can then be solved for a and b. The solution is

$$\begin{aligned}
a &= A \exp\left[-(i/\hbar)\int(a|V|b)dt\right. \\
&\quad \left. +B \exp\left[(i/\hbar)\int(a|V|b)dt\right]\right. \\
b &= A \exp\left[-(i/\hbar)\int(a|V|b)dt\right. \\
&\quad \left. -B \exp\left[(i/\hbar)\int(a|V|b)dt\right]\right.,
\end{aligned} \tag{4a}$$

where $|A|^2+|B|^2=\frac{1}{2}$.

According to the treatment given by Furssow and Wlassow the motion of the atoms is fixed and no account is taken of changes in it. The initial conditions are made such that at the initial time the atom 1 is excited and atom 2 is not. This condition requires as initial values, $a=0$, $b=1$, so the above solutions become

$$\begin{aligned}
a &= -i \sin (1/\hbar)\int_{t_0}^{t}(a|V|b)dt, \\
b &= \cos (1/\hbar)\int_{t_0}^{t}(a|V|b)dt.
\end{aligned} \tag{5}$$

Since the nuclear motion is to be described classically, the functions representing it will consist of a compact wave packet for each atom, and the distance between the atoms will have a precise value which is a function of the time. The matrix elements will be functions of R and will have the form

$$(a|V|b) = \alpha/R^3 = \alpha/(p^2+v^2t^2)^{\frac{3}{2}}, \tag{6}$$

where α is a constant which depends upon the electronic functions, p is the distance of closest approach, and v is the relative velocity of the atoms. The zero of time is taken at the time of closest approach. This expression can be integrated so that

$$\begin{aligned}
a &= -i \sin\left\{(\alpha t/\hbar p^2 v)/(p^2/v^2+t^2)^{\frac{1}{2}}\right. \\
&\quad \left. -(\alpha t_0/\hbar p^2 v)/(p^2/v^2+t_0^2)^{\frac{1}{2}}\right\}, \\
b &= \cos\left\{(\alpha t/\hbar p^2 v)/(p^2/v^2+t^2)^{\frac{1}{2}}\right. \\
&\quad \left. -(\alpha t_0/\hbar p^2 v)/(p^2/v^2+t_0^2)^{\frac{1}{2}}\right\}.
\end{aligned} \tag{7}$$

Most of the change in these functions occurs near the time $t=0$, and the whole change in phase between $t=-\infty$ and $t=+\infty$ is $2\alpha/\hbar p^2 v$. Hence the effect of a collision at the velocity v and the distance of closest approach p is to make the coefficients

$$a = -i \sin (2\alpha/\hbar p^2 v), \quad b = \cos (2\alpha/\hbar p^2 v). \tag{8}$$

This represents a probability of $\sin^2 (2\alpha/\hbar p^2 v)$ that the energy of excitation has been transferred to atom 2 during the collision. Since the collisions occur frequently during the lifetime of an excited state the coefficient b may be approximated by a decreasing exponential, $e^{-\mu t}$, which, substituted into the equation for c, gives the familiar resonance curve for the probability of radiation.

On the other hand one may insert the general solution, Eq. (4a), into the equation for c. This gives c as a function of ν by a Fourier analysis of $\exp\left\{(i/\hbar)\int(a|V|b)dt-(i/\hbar)(E_1-E_0)t\right\}$. The evaluation of this integral is just the method of Weisskopf and Lenz. Instead of treating the exact analysis, it is possible to consider the phase as changing abruptly by the amount $(2\alpha/\hbar p^2 v)$ at each collision.

In neither of these approximations is the mechanism of the conservation of energy obvious, because of the assumption of constraints which cause the atoms to move past each other in straight lines at uniform velocity. To give a treatment in which this motion is allowed to change under the influence of the atomic interaction it is necessary to make the functions $v_1(R, t)$ and $v_0(R, t)$ characteristic of the energy of translation. When this is done the matrix elements $(a|V|b)$ are constants whose magnitude depends upon the translational energy but which vanish as the relative duration of the collision decreases. When the corresponding values for a

and b are inserted in the equation for c, the integral shows a secular increase of c only when $(E_1+T_1-E_0-T_0)=h\nu$. T_1 and T_0 are the kinetic energies before and after the impact respectively. The square of the absolute value of c will be proportional to $|(c|S|a)|^2$ which will depend upon ν through the difference (T_1-T_0). This, as shown by Weisskopf, gives approximately the same distribution of radiation as the previous methods.

The conclusion is that all three methods are essentially equivalent, and that one is to be preferred over another only if the computation seems simpler or if it can be shown that the approximation to the exact solution is better. Thus far no outstanding superiority of one method over the others seems to have been demonstrated, but for the following treatment of multiplets the point of view of Furssow and Wlassow has been adopted.

Broadening of the Members of a Multiplet

To take into account the multiplet structure consider the ground state to have the inner quantum number J and the excited state to have the inner quantum number J'. When the atoms are held at a fixed distance apart there are $2(2J+1)(2J'+1)$ stationary states of the dipole interaction energy. Each of these states represents a definite value of the angular momentum about the line connecting the atoms. Positive and negative values of this angular momentum have the same value of the energy. Half of the states are symmetric and half of them are antisymmetric in the two atoms, and the corresponding symmetric and antisymmetric states have energies of equal magnitude but of opposite sign. Furthermore, in each class of states the sum of the interaction energies is zero.

Let the $2(2J+1)(2J'+1)$ states be designated as follows

$$\begin{aligned}\psi(p, M, M')&=u(\alpha', J', M'; 1)\\ &\quad\times u(\alpha, J, M; 2)v_1(R, t)e^{-(i/\hbar)E_1t}\\ \psi(q, M, M')&=u(\alpha', J', M'; 2)\\ &\quad\times u(\alpha, J, M; 1)v_1(R, t)e^{-(i/\hbar)E_1t}.\end{aligned}\tag{9}$$

The quantum numbers α', J', M' represent the excited state and α, J, M, the ground state. The designations p and q refer respectively to states in which the excitation energy is in atom 1 and to those in which it is in atom 2. These are not the symmetric and antisymmetric states which are characteristic of the interaction energy, but are selected to conform to the treatment of Furssow and Wlassow. If the interaction with the radiation field is neglected, as above, the equations to be solved are

$$\begin{aligned}\dot a(p, M, M')&=-(i/\hbar)\sum_{\mu,\mu'}\\ &\quad\times(p, M, M'|\,V|\,q, \mu, \mu')a(q, \mu, \mu')\\ \dot a(q, M, M')&=-(i/\hbar)\sum_{\mu,\mu'}\\ &\quad\times(q, M, M'|\,V|\,p, \mu, \mu')a(p, \mu, \mu').\end{aligned}\tag{10}$$

If the assumed motion of the atoms is such that the nuclei move past each other in straight lines and that the angular momentum around the line connecting the atoms remains constant, the equations can be solved. This involves the application to each atom of a torque as well as a force. To find the solution let

$$\begin{aligned}a(p, M, M')&=A(p, M, M')e^{i\omega(t)},\\ a(q, M, M')&=A(q, M, M')e^{i\omega(t)}.\end{aligned}\tag{11}$$

Substitution into Eqs. (10) gives a set of simultaneous equations for the A's whose coefficients contain the $\dot\omega$. The values of $\dot\omega$ which permit the solution of the equations are just $(i/\hbar)$ times the characteristic values of the interaction energy. Let these be α_j/R^3. The solution will then have the form

$$\begin{aligned}a(p, M, M')&=\sum_j A_j(p, M, M')\\ &\quad\times\exp\left[(i\alpha_j/\hbar)\int dt/R^3\right].\end{aligned}\tag{12}$$

In accord with the method of Furssow and Wlassow all of the $a(q, M, M')$ must be set equal to zero at the initial time. After a single collision

$$\begin{aligned}a(p, M, M')&=\sum_j A_j\cos(2\alpha_j/hp^2v),\\ a(q, M, M')&=i\sum_j A_j\sin(2\alpha_j/hp^2v),\end{aligned}\tag{13}$$

where the sum is taken over only those values of j which represent either the symmetric or the antisymmetric states. This solution assumes a

definite constant orientation of the angular momentum vectors with reference to the line connecting the atoms and this orientation determines the relative values of the coefficients A_j. To include all possibilities it is necessary to average over all sets of initial and final states. This may be done by considering that there are $(2J+1)(2J'+1)$ states in which the system might be before collision, and the same number into which it might be thrown by the collision. The transition probability to many of these is zero, but the sum of the transition probabilities must be divided by the total number of possibilities. This leads to a resultant transition probability of

$$P=(1/(2J+1)^2(2J'+1)^2)\sum_j \sin^2(2\alpha_j/\hbar p^2 v), \quad (14)$$

that the excitation energy has been transferred to the second atom. Collisions closer than those for which the probability becomes equal to unity can be regarded as collisions in the sense of the Lorentz. To determine the value of this distance, p_0, the sine may be replaced by its argument. Then

$$\frac{4}{\hbar^2 p_0^4 v^2 (2J+1)^2(2J'+1)^2}\sum_j \alpha_j^2 = \frac{8S^2(J,J')}{3\hbar^2 p_0^4 v^2(2J+1)^2(2J'+1)^2}, \quad (15)$$

where $S(J, J')$ is the strength of the line as defined by Condon and Shortly.[6] The half-width at half-maximum due to collisions of this kind is then

$$\mu_1 = N\pi p_0^2 v = (8/3)^{\frac{1}{3}}\pi S(J,J')/\hbar(2J+1)(2J'+1) \quad (16)$$

in angular units.

For the more distant collisions the contributions of a large number of impacts must be added together to get the probable rate of loss of energy. Let this rate be $2\mu_2$.

$$2\mu_2 = \frac{8NvS^2(J,J')2\pi}{3\hbar^2 v^2(2J+1)^2(2J'+1)^2}\int_{p_0}^{\infty}\frac{dp}{p^3} = \left(\frac{2}{3}\right)^{\frac{1}{3}}\frac{2\pi S(J,J')N}{\hbar(2J+1)(2J'+1)}. \quad (17)$$

The rate at which the probability amplitude decreases is just half of this so that the total half-width at half-maximum in angular measure is

$$\mu=\mu_1+\mu_2=\frac{6^{\frac{1}{3}}\pi S(J,J')}{\hbar(2J+1)(2J'+1)}N. \quad (18)$$

From the definition of $S(J, J')$ it follows that this width is the same for all members of a Russell-Saunders multiplet.

Comparison with Experiment

The work of Hughes and Lloyd[2] showed the equality of the broadening of the two members of the first doublet of the principal series of potassium, and also showed the proportionality of the width with N. The measurement of the width showed that

$$\Delta\nu_{\text{obs}}=16\times10^{-8}N\text{cm}^{-1}.$$

For this doublet the line strength is[7]

$$S(\tfrac{1}{2},\tfrac{1}{2})=e^2h/4\pi^2 m\nu=1.09\times10^{-34}.$$

This leads to a theoretical half-width at half-maximum of

$$\Delta\nu_{\text{calc}}=6^{\frac{1}{3}}Ne^2/16\pi m\nu=3.16\times10^{-8}N\text{cm}^{-1}.$$

A possible explanation of this discrepancy of a factor of five lies in the uncertainty with respect to the vapor pressure of potassium as a function of temperature. The value used by Lloyd was extrapolated by means of a formula from a region where there was already considerable divergence among different observations. It is hard to believe that the theoretical approximations used could account for as much as a factor of two.

It is an honor and a privilege to be permitted at this time to pay my respects publicly to Professor Sommerfeld on the occasion of his seventieth birthday. His influence on physics and physicists will long endure.

[6] Condon and Shortley, *The Theory of Atomic Spectra*, p. 98.
The fact that $\sum_j \alpha_j^2 = \frac{2}{3}S^2$ can be derived by writing down the expression for the matrix elements of the dipole interaction energy and summing the squares of the absolute values. This gives the diagonal sum of the matrix of the square of this energy.

[7] This assumes the f value of the first doublet to be unity which is correct to about one percent.

Rigorous Theory of Interaction Between Nuclear Particles

E. C. G. STUECKELBERG
Institut de Physique de l'Université de Genève, Geneva, Switzerland
(Received September 21, 1938)

The static interaction between elementary particles can be introduced in the Hamiltonian by way of a contact transformation. In classical theory no serious difficulties arise. However, in quantum theory of radiation divergent terms appear.

1. The Field Equations for Radiation and Matter

WE use the term radiation for a set of real fields Q_r satisfying

$$((\partial^2/\partial x_k \partial x_k - l^2)Q_r = -4\pi(J_r - \partial S_{rk}/\partial x_k). \quad (1.1)$$

In this chapter a double tensor index k implies summation from 1 to 4. There is no classical way to treat the spin and charge variables of elementary particles.[1] Therefore we represent matter by a set of complex wave fields φ_ν. The quantities J_r and S_{rk} occurring in (1.1) are formed from matrices j_r and s_{rk}:

$$J_r = \sum\sum \varphi_\nu^* j_{r\nu\nu'} \varphi_{\nu'} = \varphi^* j_r \varphi; \quad S_{rk} = \cdots \quad (1.2)$$

φ satisfies

$$ih\dot{\varphi} = (\kappa - \sum(Q_r j_r + (\partial Q_r/\partial x_k)s_{rk}))\varphi. \quad (1.3)$$

κ is the Dirac operator

$$\kappa = -ihc(\boldsymbol{\alpha}\cdot\text{grad}) + mc^2\beta. \quad (1.4)$$

(1.1) and (1.3) apply to both electrodynamic and nuclear forces.

We write Φ_k for Q_r identifying r with the tensor index k. The matrices are given in Table I. ϵ and μ commute with each other and with $\boldsymbol{\alpha}$ and β. From (1.3) follows $\partial J_k/\partial x_k = 0$. Therefore Φ_k can be submitted to the initial condition $\partial\Phi_k/\partial x_k = 0$. If we define

$$F_{kk'} = \partial\Phi_{k'}/\partial x_k - \partial\Phi_k/\partial x_{k'} + 4\pi S_{kk'}, \quad (1.5)$$

equation (1.1) takes the form of Maxwell's equations

$$\partial F_{kk'}/\partial x_k - l^2\Phi_{k'} = -4\pi J_{k'} \quad (1.6)$$

if $l=0$. Eq. (1.3) takes the form of Dirac's equation for a particle with inherent magnetic moment μ (proton ?).[2] A term $-2\pi S_{kk'}s_{kk'}\varphi$ is added to (1.3). The last part of (1.3) now takes the form

$$(\cdots - \Phi_k j_k + \tfrac{1}{2}F_{kk'}s_{kk'})\varphi. \quad (1.7)$$

We obtain the equations of nuclear interaction proposed in scalar form by Yukawa[3] and generalized into tensor form by different authors,[4] if r is a double index i, k and if k goes from 1 to 5. Let $\Phi_k{}^i (k=1 \text{ to } 4)$ be defined by

$$\Phi_k{}^i = Q_k{}^i + l^{-1}\partial Q_5{}^i/\partial x_k. \quad (1.8)$$

The matrices are given in Table II. The initial condition

$$\partial Q_k{}^i/\partial x_k + lQ_5{}^i = \partial(\Phi_k{}^i - 4\pi l^{-2}J_k{}^i)/\partial x_k = 0 \quad (1.9)$$

is compatible with Eq. (1.1). If the $F_{kk'}{}^i$ are again defined by (1.5), the Eqs. (1.6) with $l \neq 0$ are valid. Combining two real fields with a complex field the usual equations for charged and uncharged Yukawa particles result.

2. Hamiltonian Functional and Functional Contact Transformation

Let the functional derivative $\delta F/\delta G$ of a functional $F[G(x)]$ be defined from the variation $\delta F = \int dx(\delta F/\delta G)\delta G(x)$ if δG is an arbitrary but small variation of the function $G(x)$. x represents the ensemble of space coordinates. To eliminate

[1] The quantum theoretical treatment has been considered in a previous note: Stueckelberg, Comptes rendus **207**, 387 (1938).

[2] Pauli, *Handbuch der Physik* Vol. **24**, 24/1 (2nd ed. 1933), p. 233.

[3] Yukawa, Proc. Phys. Math. Soc. Japan **17**, 48 (1935).

[4] Stueckelberg, Helv. Phys. Acta **11**, 225, 229 (1938). Yukawa, Sakata and Taketani, Proc. Phys. Soc. Japan **20**, No. 4 (1938). Kemmer, Proc. Roy. Soc. **166**, 127 (1938). Bhabha, Proc. Roy. Soc. **166**, 501 (1938).

TABLE I. *Electrodynamics.**

$$W=\pi S_{kk'}S_{kk'} \qquad \sigma_3=i\beta\alpha_1\alpha_2$$
$$H^{F'}=\cdots+\frac{\epsilon^{\nu}\epsilon^{\nu'}}{r_{\nu\nu'}}-\mu^{\nu}\mu^{\nu'}(\boldsymbol{\sigma}^{\nu}\cdot\mathrm{grad})(\boldsymbol{\sigma}^{\nu'}\cdot\mathrm{grad})\frac{1}{r_{\nu\nu'}}$$
$$r_{\nu\nu'}=|\mathbf{q}^{\nu}-\mathbf{q}^{\nu'}|.$$

$r=k$	STATIC	DYNAMIC
1, 2, 3	0	$j_k=\epsilon\alpha_k$
	$s_{kk'}=i\mu\beta\alpha_k\alpha_{k'}$	0
	0	$ic\kappa_k=s_{k4}=-\mu\beta\alpha_k$
4	$j_4=i\epsilon$	0
	0	$s_{kk'}=\mu\beta\alpha_{k'}$
	0	0

* Scalar products of operators symbolize summation over index k (for σ_k) and index i (for ϵ^i and μ^i) k, k' takes the values 1 to 3.

boundary effects all functions are submitted to a periodicity condition.

Let $H(P, Q, \varphi^*, \varphi)$ be a functional of the four sets of functions. The extremum condition

$$\delta\int dt(H-\sum P_r\dot{Q}_r-ih\sum\varphi_\nu^*\dot{\varphi}_\nu)=0 \qquad (2.1)$$

leads to the canonical equations

$$\dot{Q}_r=\delta H/\delta P_r; \quad -\dot{P}_r=\delta H/\delta Q_r;$$
$$ih\dot{\varphi}_\nu=\delta H/\delta\varphi_\nu^*; \quad -ih\dot{\varphi}_\nu^*=\delta H/\delta\varphi_\nu. \qquad (2.2)$$

Subtracting the total differential $\dot{U}$ of a functional $U(Q, P', \varphi'^*, \varphi)$ from the integrand of (2.1) the extremum condition in the new (primed) variables keeps the form (2.2), if the relation between primed and unprimed variables is the functional contact transformation

$$Q_r'=\delta H/\delta P_r'; \quad P_r=\delta H/\delta Q_r;$$
$$ih\varphi'=\delta H/\delta\varphi'^*; \quad ih\varphi^*=\delta H/\delta\varphi. \qquad (2.3)$$

U has to be chosen in such a way as to make φ'^* the conjugate complex of φ'. If U has the form

$$U=\int dx(\Sigma P_r'Q_r+ih\varphi'^*u\varphi) \qquad (2.4)$$

(the matrix u depends but on P'), the relations

$$\varphi^*=\varphi'^*u; \quad \varphi=u^+\varphi' \qquad (2.5)$$

hold, if $u^+u=1$. (The superscript $^+$ denotes the Hermitian conjugate matrix.)

To set up the Hamiltonian we define the differential operator Z

$$4\pi Z=l^2-\Delta. \qquad (2.6)$$

The inverse operators Z^{-n} for $n=1, 2$ can be defined as

$$Z^{-n}f(x)=\int dy v_n(|y-x|)f(y) \qquad (2.7)$$

with

$$v_1(r)=r^{-1}e^{-lr}; \quad v_2(r)=2\pi l^{-1}e^{-lr}. \qquad (2.8)$$

If we set

$$H=H^F+H^M+V \qquad (2.9)$$

with

$$2H^F=\int dx\sum(Q_rZQ_r+fP_r^2); \quad f=4\pi c^2$$
$$H^M=\int dx(\varphi^*\kappa\varphi+W), \qquad (2.10)$$
$$V=\int dx\sum(-Q_r(J_r-\partial S_{rk}/\partial x_k)-fP_rK_r), \qquad (2.11)$$
$$icK_r=-S_{r4}.$$

Contrary to Section 1 all summations over double indices k go from 1 to 3 in this section. W contains biquadratic terms in φ. Its value is given in Tables I and II. Translated into particle theory W expresses a direct interaction between particles (=charge elements) whose potential has the form of a δ-function.

The contact transformation (2.3) leads to (2.5) and to

$$P_r=P_r'; \quad Q_r=Q_r'+Z^{-1}(J_r'-\partial S_{rk}'/\partial x_k) \qquad (2.12)$$

if the matrix u satisfies

$$\partial u/\partial(iZ^{-1}P_r'/h)=j_ru;$$
$$\partial u/\partial(iZ^{-1}\partial P_r'/\partial x_k)=s_{rk}u \qquad (2.13)$$

k goes again from 1 to 3. J_r' is formed from the φ' with the same matrix j_r as in (1.2). A second transformation functional $\bar{U}$

$$\bar{U}=\int dx(\sum P_r'Q_r''+ih\varphi^{*\prime}\bar{u}^+\varphi'') \qquad (2.14)$$

TABLE II. *Nuclear forces.**

$$W=\sum_i(\pi S_{kk'}{}^iS_{kk'}{}^i+\frac{f}{2}(K_5^i)^2)$$
$$H^{F'}=\cdots\frac{(\boldsymbol{\varepsilon}^{\nu}\cdot\boldsymbol{\varepsilon}^{\nu'})}{r_{\nu\nu'}}e^{-lr_{\nu\nu'}}+(\boldsymbol{\mu}^{\nu}\cdot\boldsymbol{\mu}^{\nu'})((\boldsymbol{\sigma}^{\nu}\cdot\boldsymbol{\sigma}^{\nu'})l^2$$
$$-(\boldsymbol{\sigma}^{\nu}\cdot\mathrm{grad})(\boldsymbol{\sigma}^{\nu'}\cdot\mathrm{grad}))\frac{e^{-lr_{\nu\nu'}}}{r_{\nu\nu'}}\cdot$$

$r=^ik$	STATIC	DYNAMIC
1, 2, 3	0	$j_k{}^i=\epsilon^i\alpha_k$
	$s_{kk'}{}^i=i\mu^i\beta\alpha_k\alpha_{k'}$	0
	0	$ic\kappa_k{}^i=s_{k4}{}^i=-\mu^i\beta\alpha_k$
4	$j_4{}^i=i\epsilon^i$	0
	0	$s_{4k'}{}^i=\mu^i\beta\alpha_{k'}$
	0	0
5	0	0
	0	$s_{kk'}{}^i=-l^{-1}j_{k'}{}^i$
	$ic\kappa_5{}^i=s_{54}{}^i=-l^{-1}j_4{}^i$	0

with $\bar{u}^+$ depending on Q'' in such a way as to satisfy

$$\partial\bar{u}^+/\partial(iQ_\tau''/h)=\bar{u}^+\kappa_\tau \tag{2.15}$$

transforms φ in an analogous way to (2.5). It leaves Q invariant and leads to

$$P_\tau'=P_\tau''+K_\tau''; \quad K_\tau''=\varphi''^*\kappa_\tau\varphi''. \tag{2.16}$$

Omitting the primes from the variables, the transformed Hamiltonian takes the form

$$\begin{aligned} &H^{F1}=H^F,\\ &H^{M1}=H^M-\tfrac{1}{2}\int dx\sum((J_\tau-\partial S_{\tau k}/\partial x_k)Z^{-1}\\ &\qquad\times(J_\tau-\partial S_{\tau k}/\partial x_k)+fK_\tau^2),\\ &V'=\int dx\varphi^*[u,\kappa]u^+\varphi. \end{aligned} \tag{2.17}$$

u in (2.17) signifies the matrix $\bar{u}u$. $[a,b]$ is the commutator between two operators a and b.

In Tables I and II we distinguish between *static* and *dynamic* matrices. To them correspond in V (2.10) static and dynamic interaction terms between matter and radiation. If ν represents the frequency of φ, the dynamic terms are smaller than the static ones by a factor $h\nu/mc^2$. In a particle representation the dynamic terms correspond to relativistic effects. If the transformed Hamiltonian contains in V' but dynamic terms, the solutions of the equations

$$P_\tau=Q_\tau=0; \quad ih\dot{\varphi}=\delta H^{M1}/\delta\varphi^* \tag{2.18}$$

are a zeroth approximation in $h\nu/mc^2$ (=non-relativistic approximation for small particle velocities).

Let the matrices u and $\bar{u}$ depend only on the static terms of Tables I and II. The transformed V' consists now of the old dynamic terms plus new dynamic terms of the form (2.17).

In H^{M1} we find the static terms of the old W plus new static interaction terms matter-matter. In a particle representation one sets

$$J_\tau(x)=\sum_\nu j_\tau{}^\nu\delta(q^\nu-x), \tag{2.19}$$

q^ν and $j_\tau{}^\nu$ being space coordinates and matrices referring to the νth particle. Substitution of (2.19) into (2.17) gives the well-known interaction formulas (in Tables I and II).

However a serious difficulty arises: If the matrices j_τ and $s_{\tau k}$ (referring to static terms only) do not commute, Eqs. (2.13) are not integrable. Choosing

$$\begin{aligned} &u=\exp\left((i/h)\sum(j_\tau Z^{-1}P_\tau'+s_{\tau k}Z^{-1}\partial P_\tau'/\partial x_k)\right)\\ &\bar{u}^+=\exp\left((i/h)\sum\kappa_\tau Q_\tau''\right) \end{aligned} \tag{2.20}$$

(the summation is extended over static terms only) we obtain additional terms in (2.12) and (2.16) due to this noncommutability. Other similar additional terms arise from $uj_\tau u^+=j_\tau+\cdots$. If the matrices j_τ etc. are all proportional to a small quantity (=charge e) these additional terms are a series in e beginning with e^2. Only for the Dirac electron ($\mu=0$), which has but one static term j_4, these higher terms disappear.

If we define the matrices proportional to e^2 as:

$$\begin{aligned} &e^2a_{\tau,\tau'}=(i/h)[j_\tau,j_{\tau'}];\\ &e^2a_{\tau k,\tau'}=(i/h)[s_{\tau k},j_{\tau'}]; \quad e^2a_{\tau k,\tau' k'}=\cdots,\\ &e^2b_{\tau,\tau'}=(i/h)[\kappa_\tau,\kappa_{\tau'}],\\ &e^2c_{\tau,\tau'}=(i/h)[j_\tau,\kappa_{\tau'}]; \quad e^2c_{\tau k,\tau'}=\cdots, \end{aligned} \tag{2.21}$$

V' contains the following additional static terms:

$$\begin{aligned} V'=&\,e\times\text{dynamic terms}\\ &+e^2\times\tfrac{1}{2}\sum\{-Q_\tau(A_{s,\tau}Z^{-1}P_s\\ &\quad+A_{sk,\tau}Z^{-1}\partial P_s/\partial x_k)\\ &\quad-(\partial Q_\tau/\partial x_l)(A_{s,\tau l}\cdots)-P_\tau B_{s,\tau}Q_s\\ &\quad-2fP_\tau(C_{s,\tau}Z^{-1}P_s+C_{sk,\tau}\cdots)\}\\ &+e^3\times\cdots+\cdots,\\ &\quad A_{s,\tau}=\varphi^*a_{s,\tau}\varphi; \quad A_{sk,\tau}=\cdots. \end{aligned} \tag{2.22}$$

The transformed radiation field equations are, instead of (1.1), of the form

$$\begin{aligned} (\sum_1^4\partial^2/x\partial_k\partial x_k-l^2)Q_\tau=&\,e\times\text{dynamic terms}\\ &+e^2\times\text{terms linear in } Q_s \text{ and } P_s\\ &+e^3\times\cdots+\cdots. \end{aligned} \tag{2.23}$$

Neglecting dynamic effects and taking only terms up to e^2 into account, we find the solution $P_\tau=Q_\tau$ is still possible. In this approximation matter still satisfies Eq. (2.18) which takes into account only static interaction proportional to e^2. The terms of third order in V' (2.22) are again linear in Q and P. A further contact transforma-

tion leads, therefore, to new static interaction terms matter-matter proportional to e^6 etc.

After all these transformations (2.23) will contain only dynamic inhomogeneities proportional to the odd powers of e and terms which vanish with vanishing fields.

3. Quantum Theory

The canonical equations (2.2) result from H if P_r and Q_r satisfy the usual commutation rules

$$[P_r(x), Q_{r'}(x')]=(i/h)\delta_{rr'}\delta(x-x'). \quad (3.1)$$

The φ and φ^* satisfy the commutation or anti-commutation rules of Bose or Fermi statistics.

The contact transformation U transforms a physical variable (operator) F into F' if

$$F'=UFU^+; \quad UU^+=1. \quad (3.2)$$

Eqs. (2.12) result from (3.2) if U satisfies

$$\delta U/\delta(iP_r'/h)=Z^{-1}(J_r'-\partial S_{rk}'/\partial x_k)\cdot U \quad (3.3)$$

(an analogous relation holds for $\bar{U}$). The same difficulties (=nonintegrability) as in classical theory occur if the J_r, K_r and S_{rk} do not commute. Setting

$$U=\exp\left((i/h)\int dx\sum(J_r'Z^{-1}P_r' + S_{rk}'Z^{-1}\partial P_r'/\partial x_k\right) \quad (3.4)$$

and the analogous expression for $\bar{U}$, we arrive at exactly the same H^{F_1}, H^{M_1} and V' as in (2.17).

After effecting all the contact transformations mentioned at the end of Section 2, we arrive at a Hamiltonian which contains, besides the static interaction matter-matter (terms in e^2, e^6, etc.) and the dynamic terms, the expressions (2.22) and similar higher terms. To eliminate them in a nonrelativistic approximation, it is not possible to set $Q_r=P_r=0$, because the two variables do not commute. Applying perturbation theory similar to the old theory of nuclear interaction with the electron-neutrino field, the additional terms (2.22) give a static interaction proportional to e^4. Exactly as in neutrino-electron field theory their dependence on the distance r leads to diverging integrals[5] which can be made convergent. For small values of r these expressions go to ∞ proportional to a large inverse power of r.

Since these terms are established by perturbation theory and not by contact transformation, they are to be considered only as first-order perturbations on the energy levels calculated from the rigorous static terms. Because of the high negative power of r, these first-order perturbations are diverging, if the ψ-function of the mutual orbit of two particles does not vanish for $r=0$.

It is now proposed to calculate the ground level of the deuterium nucleus first from the rigorous classical terms (e^2, e^6 etc.) and to discuss afterwards the influence of the e^4-terms. (Cutting off the integrals for an appropriate fundamental length.)

In contributing this note the writer wishes to express his admiration for Professor Sommerfeld's research work in quantum theory and his gratitude for the stimulating periods during which the writer had the great privilege to work under Professor Sommerfeld's supervision.

[5] Bethe and Bacher, Rev. Mod. Phys. **8**, 202 (1936).

The Force of Excitation in the Theory of Dispersion

P. P. EWALD
Crystallographic Laboratory, Cambridge, England
(Received October 17, 1938)

IN the classical theory of the propagation of an optical wave through a transparent body, each individual atom is replaced by a dipole of appropriate polarizability α. The electric moment p acquired by the atom under the influence of the optical field passing over it is $p=\alpha E$, where E is the momentary and local value of the electric force of the optical field. The correct determination of this field is one of the chief difficulties in the theory of optical propagation. The validity of the Lorentz-Lorenz formula, the intensity of sideways scattering in amorphous media, the amount of double refraction and of optical rotatory power in crystals, are among the properties the theory of which is directly influenced by the expression obtained for the optical field acting upon a particular atom.

For a body represented by a perfect crystal lattice, part of the difficulties of calculation are removed because the problem contains no statistical bias.[1] If we restrict our attention to the propagation of the optical field in the interior of the body, we may take the body to be unlimited and we need not worry about an external incident wave by which the optical field, now in the interior, has been started. The field of excitation acting on a particular atom would then seem to consist of the sum of the electric fields radiated to the atom in question by all the other atoms of the crystal. Or, in other words, if we form the total optical field E_{tot} by summing the contributions of *all* atoms, we arrive at the field of excitation E_{exc} by subtracting from E_{tot} the contribution of the particular atom which is to be acted upon. At the place of the atom, both its own dipole field and E_{tot} will become infinite, but E_{exc} remains finite.

The electric force E_{exc} is, however, not the correct force to be inserted into the equation of motion $p=\alpha E$ of the atom. The reaction of the dipole's own field on its movement is missing. This reaction is necessary in order to comply with the conservation of energy, for if only external fields acted on a dipole, a dipole once started to vibrate by a momentary impulse of external origin could not come to rest again because, with no external field on, there would be no force acting to stop the vibrations. Part of the dipole's own field of radiation has thus to be included in the field used to determine the dipole moment of the atom. Instead of this, α may be taken complex, or a term due to radiation damping included in the equation of motion of the dipole. Only if this is done, will the correct force acting on the dipole be represented by E_{exc}. In the case of a crystal, E_{exc} contains terms which cancel against the contribution of the dipole's own field or which counteract the phase difference introduced by the complexity of α, and there is no radiation damping perceptible in the crystal—a result first announced by H. A. Lorentz.[2]

As long as the atoms are considered to form independent units in the structure of the crystal, there is thus no serious difficulty in establishing the equation of motion for them. According to the wave-mechanical concept of the crystal, this independence ceases and, in a sense, each atom can be said to extend through the whole domain of the crystal. What then becomes of the difference between the total optical field and the field of excitation? To answer this question we have to consider the two possible ways of approach to a solution of the problem of propagation of an optical field through the crystal.

The more fundamental of the two ways would be to treat the whole crystal as *one* system having stationary states which may, for instance,

[1] This advantage was clearly recognized by A. Sommerfeld in 1909 when he proposed to me the subject for my thesis, *Dispersion und Doppelbrechung vom Elektronengittern* (*Kristallen*), (München, 1912).

[2] The field of excitation has been thoroughly discussed by me in a paper appearing shortly in Annales de l'Institut H. Poincaré.

be formed by linear combinations of the states of the individual atoms. The states of the crystal ordinarily investigated are those obtained in the absence of external forces. They are optically ineffective, i.e., the resultant dipole moment of each lattice cell is zero, except for effects of metallic conductivity. If a light wave is present, we could consider it, in analogy to the atomic case, as a perturbation acting on the wave functions of the crystal. The perturbed wave function thus obtained gives a nonvanishing dipole moment in each lattice cell; it is "optically effective." The dipole moment travels through the crystal with the same speed and wave-length as the optical field. We must then consider the influence of this dipole wave back on the optical field, and we must make the wave function of the crystal and the optical field mutually consistent. Thus, actually, the wave function of the crystal in its optically effective state cannot be determined by a straightforward application of perturbation theory but it must be found, simultaneously with the optical field, from a zero-order Hamiltonian which already contains the interaction between crystal and optical field. The further approximations serve to determine accurately the influence which the optical field exerts on the states of the atoms in the crystal as compared with the states prevailing in the optically ineffective case. The absorption would, for example, be a result of a further approximation.

So far as the author is aware, there is but one paper aiming at this kind of theory, that of G. Wentzel.[3] No distinction is being made between the potential of the total optical field and that of the field of excitation. It is therefore not to be expected that the influence of the anisotropy of the atomic arrangement would be taken into account correctly if it were possible to carry through the theory to numerical calculation.

The less fundamental, but probably more useful, form of a wave-mechanical theory of wave propagation would be one modeled very much on the lines of the classical theory. The atoms or the contents of the cells would be considered as independent units and the only place where wave-mechanics enters would be in the calculation of the polarizability of these units. This polarizability might be different from that for the unit not surrounded by the other groups because of the deformation of the ground state of the atoms when they are being introduced into the crystal lattice. (Compare with Bethe's paper on the splitting of term values for atoms introduced into a crystal.) If the coupling between the atoms is weak, the usual result will be obtained that each atom scatters as if it consisted of a system of virtual resonators whose frequencies correspond to the transitions of the atom from its ground state to various excited states. The field of the optical wave would be considered as an external field and would not, as in the more fundamental type of theory, form part of the whole system. For the field of excitation the question then arises whether to consider each of the virtual resonators of the atom as an independent unit or whether to keep those belonging to one atom together and to subtract their joint field from the total optical field. The latter course would evidently have to be followed, since it is the entire field of the atom that has to be taken away provided that the radiation damping is taken care of explicitly.

Summing up, we may state that there are still difficulties of a fundamental nature to be overcome for a wave-mechanical theory of optical wave propagation of a sufficient degree of perfection to allow a safe numerical calculation of the optical properties of crystals.

[3] G. Wentzel, Helv. Phys. Acta 6, 89 (1933).

The Diffraction of Waves by Ribbons and by Slits

PHILIP M. MORSE AND PEARL J. RUBENSTEIN
Eastman Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received September 29, 1938)

The exact solution of the diffraction problem in elliptic cylinder coordinates is computed by the use of newly computed tables of Mathieu functions. Curves are given for the distribution-in-angle and total intensity scattered, for the diffraction by a slit or the scattering by a thin ribbon, of waves whose length is the same order of magnitude as the slit or ribbon width. The curves are for different angles of incidence, for different wave-lengths, and for two boundary conditions: zero value and zero normal gradient.

THE exact solution of the problem of the diffraction of waves by a straight edge has been obtained by Sommerfeld[1] by the use of a contour integral representation. The corresponding solution[2] for the diffraction through a slit, or by a ribbon, presents certain mathematical difficulties which can, however, be overcome by separating the wave equation in elliptic cylinder coordinates and computing the values of the resulting Mathieu function solutions. The problem has been discussed previously,[3] but few numerical solutions were computed. Since a table of the needed Mathieu functions[4] has now been completed by the authors, it is of interest to discuss the properties of these functions and the nature of the solutions of the diffraction problem.

The coordinates used are given by the equations

$$\begin{aligned} x&=(d/2)\cosh\xi\cos\varphi;\\ y&=(d/2)\sinh\xi\sin\varphi;\quad z=z. \end{aligned} \tag{1}$$

We shall consider waves whose propagation vector is in the x-y plane, so that the z coordinate may be omitted from the discussion. A possible solution of the wave equation is then $G(\xi)H(\varphi)e^{-2\pi i\nu t}$ where G and H satisfy the differential equations

$$\begin{aligned} &(d^2G/d\xi^2)+(c^2\cosh^2\xi-b)G=0;\\ &(d^2H/d\varphi^2)+(b-c^2\cos^2\varphi)H=0, \end{aligned} \tag{2}$$

in which b is the separation constant and $c=(\pi d/\lambda)$, λ being the wave-length.

The periodic solutions[5] of the equation in φ are of two types: even about $\varphi=0$, and odd about $\varphi=0$. They are possible only for certain characteristic values of b. The even functions are denoted by $Se_m(c,\cos\varphi)$ and the corresponding values of b are b_m, with the sequence in m according to increasing values of b. The corresponding odd functions and characteristic values are called $So_m(c,\cos\varphi)$ and b_m'. These functions have been chosen so that

$$\begin{aligned} &Se_m(c,1)=1;\quad [(d/d\varphi)Se_m(c,\cos\varphi)]_{\varphi=0}=0;\\ &So_m(c,1)=0;\quad [(d/d\varphi)So_m(c,\cos\varphi)]_{\varphi=0}=1. \end{aligned} \tag{3}$$

The functions are orthogonal, and their normalizing factors are given by the equations

$$\int_0^{2\pi}[Se_m]^2d\varphi=N_m;\quad \int_0^{2\pi}[So_m]^2d\varphi=N_m'. \tag{4}$$

To obtain the "radial" functions dependent on ξ, we note that the equation for ξ is the same as that for φ if we set $\varphi=i\xi$. Consequently, $Se_m(c,\cosh\xi)$ or $So_m(c,\cosh\xi)$ is a solution of Eq. (2) for the same characteristic values of b_m. It is, however, more convenient to use the

[1] Sommerfeld, Math. Ann. **47**, 317 (1896); Zeits. f. Math. u. Physik **46**, 11 (1901). Calculations have also been made using parabolic cylinder functions, see Epstein, Diss. Munich (1914), and Crudelli, Nuovo Cimento **11**, 277 (1916).

[2] Schwarzschild, Math. Ann. **55**, 177 (1902).

[3] Sieger, Ann. d. Physik **27**, 626 (1908).

[4] To be published elsewhere. Preliminary mimeographed copies of the tables can be obtained from the Department of Physics, M. I. T. The tables were in part computed by means of the Bush Differential Analyzer; and the authors wish to thank Professor S. E. Caldwell and the staff of the Analyzer for their cooperation in this part of the task.

[5] See Stratton and Morse, Proc. Nat. Acad. **21**, 51 and 56 (1935).

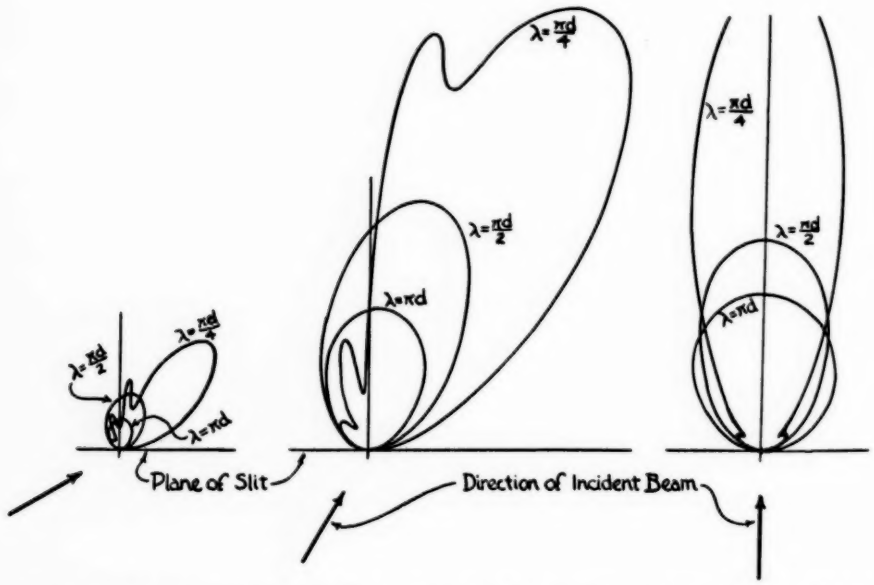

FIG. 1. Polar diagrams for distribution in angle of waves diffracted by a slit of width d (for wave function zero at boundary), or scattered by a ribbon of width d (for normal gradient zero at boundary) as function of wave-length λ and of angle of incidence of primary plane wave.

functions

$$Re_m^1(c,\rho)=[1/(2\pi)^{\frac{1}{2}}\lambda_m]Se_m(c,\rho)\underset{\rho\to\infty}{\longrightarrow}[1/(c\rho)^{\frac{1}{2}}]\cos(c\rho-(2m+1)/4\pi),$$
$$Ro_m^1(c,\rho)=[i/(2\pi)^{\frac{1}{2}}\lambda_m']So_m(c,\rho)\underset{\rho\to\infty}{\longrightarrow}[1/(c\rho)^{\frac{1}{2}}]\cos(c\rho-(2m+1)/4\pi),\qquad \rho=\cosh\xi, \tag{5}$$

where λ_m and λ_m' are constants whose values depend on the values of c. The second solutions of the equations for ξ are defined as

$$Re_m^2(c,\rho)\underset{\rho\to\infty}{\longrightarrow}[1/(c\rho)^{\frac{1}{2}}]\sin(c\rho-(2m+1)/4),$$
$$Ro_m^2(c,\rho)\underset{\rho\to\infty}{\longrightarrow}[1/(c\rho)^{\frac{1}{2}}]\sin(c\rho-(2m+1)/4). \tag{6}$$

At $\rho=1$ or $\xi=0$ (that is, along a ribbon of width d in the x-z plane, between the lines $y=0$, $x=\pm(d/2)$), the values and slopes of the functions are

$$Re_m^1(c,1)=1/(2\pi)^{\frac{1}{2}}\lambda_m;\quad Re_m^2(c,1)=-(2\pi)^{\frac{1}{2}}\mu_m;$$
$$Ro_m^1(c,1)=0;\quad Ro_m^2(c,1)=-(2\pi)^{\frac{1}{2}}\lambda_m';$$
$$[(d/d\xi)Re_m^1(c,\cosh\xi)]_{\xi=0}=0;$$
$$[(d/d\xi)Re_m^2(c,\cosh\xi)]_{\xi=0}=(2\pi)^{\frac{1}{2}}\lambda_m;$$
$$[(d/d\xi)Ro_m^1(c,\cosh\xi)]_{\xi=0}=1/(2\pi)^{\frac{1}{2}}\lambda_m';$$
$$[(d/d\xi)Ro_m^2(c,\cosh\xi)]_{\xi=0}=(2\pi)^{\frac{1}{2}}\mu_m'. \tag{7}$$

The constants λ and μ are given in the tables.[4]

It can be shown that a plane wave with propagation vector $k=(2\pi/\lambda)$ in the x-y plane and at an angle u with the x axis is given by

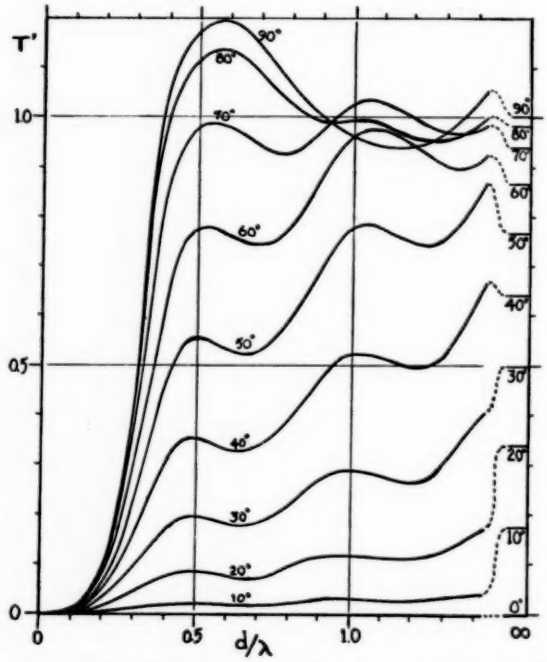

FIG. 2. Total transmission factor corresponding to the case of Fig. 1. Factor T' is the ratio between the total intensity diffracted or scattered and that required by geometrical optics for normal incidence.

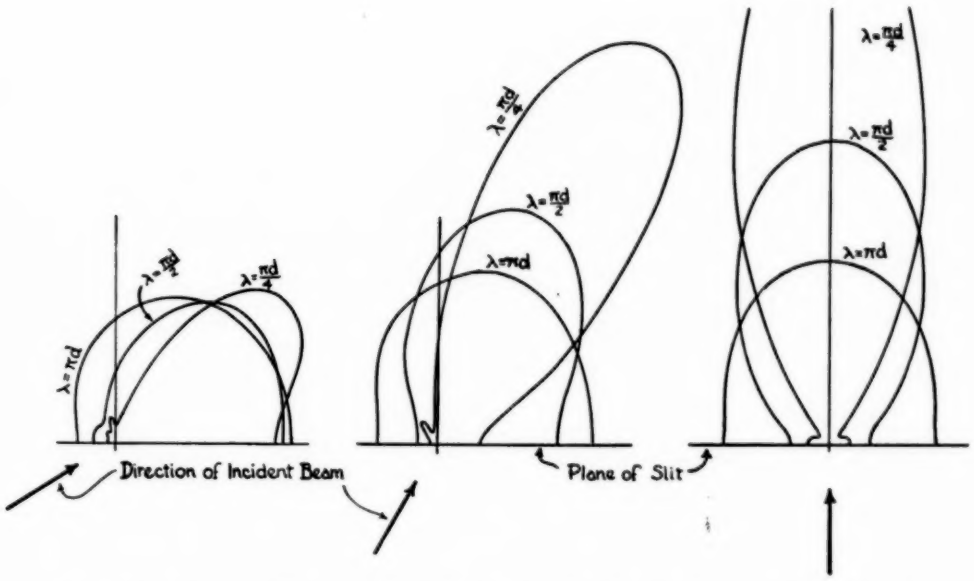

FIG. 3. Polar diagrams for waves scattered by a ribbon (for wave function zero at ribbon) or diffracted by a slit (for normal gradient zero at boundary).

the following series

$$e^{ik(x\cos u+y\sin u)}=e^{ic(\cosh\xi\cos\varphi\cos u+\sinh\xi\sin\varphi\sin u)}$$

$$=(8\pi)^{\frac{1}{2}}\sum_m i^m[(1/N_m)Se_m(c,\cos u)$$

$$\times Se_m(c,\cos\varphi)Re_m^{1}(c,\cosh\xi)$$

$$+(1/N_m')So_m(c,\cos u)$$

$$\times So_m(c,\cos\varphi)Ro_m^{1}(c,\cosh\xi)], \quad (8)$$

where $c=(kd/2)=(\pi d/\lambda)$.

By the use of the properties given in Eqs. (3)–(8), it is possible to compute the scattering from an elliptic cylinder whose axes correspond to the lines $y=0$, $x=\pm(d/2)$, and in particular from the ribbon corresponding to the limiting case of a cylinder of zero thickness, $\xi=0$. If the boundary condition is that the normal gradient of the wave function ψ is zero at the ribbon, the correct solution for a plane wave incident at an angle u to the ribbon, plus the necessary scattered wave, is

$$\psi=e^{ik(x\cos u+y\sin u)}+(8\pi)^{\frac{1}{2}}\sum_m(i^m/N_m')\sin\gamma_m'e^{i\gamma_m'}$$

$$\times So_m(c,\cos u)So_m(c,\cos\varphi)$$

$$\times[Ro_m^{1}(c,\cosh\xi)+iRo_m^{2}(c,\cosh\xi)], \quad (9)$$

where $\cot\gamma_m'=2\pi\lambda_m'\mu_m'$. The angles γ are given in the tables.[4] The intensity scattered at an angle φ to the plane of the ribbon per unit incident intensity therefore becomes $(d/r)I'$ at large distances from the ribbon, where

$$I'=(4\pi/c)\sum_{m,n}(1/N_m'N_n')\sin\gamma_m'\sin\gamma_n'$$

$$\times\cos(\gamma_n'-\gamma_m')So_m(c,\cos u)So_n(c,\cos u)$$

$$\times So_m(c,\cos\varphi)So_n(c,\cos\varphi) \quad (10)$$

and r is the distance from the center line of the ribbon. This angular distribution is the correct solution of the problem of the scattering of sound waves from a solid ribbon of width d, and also of that of the scattering of electromagnetic waves polarized so that the magnetic vector is parallel to the ribbon axis. Further investigation shows that this distribution is also correct for the diffraction of waves through a slit of width d. The boundary surface in this case is given by $\varphi=0$ and $\varphi=\pi$; and the boundary condition requiring the result (10) is that the wave function go to zero at the surface, corresponding to electromagnetic waves with their electric vector parallel to the slit axis. Eq. (10) then gives the intensity of the diffracted wave resulting when a plane wave of unit intensity is incident on the other side of the slit, directed at an angle u to the plane of the slit.

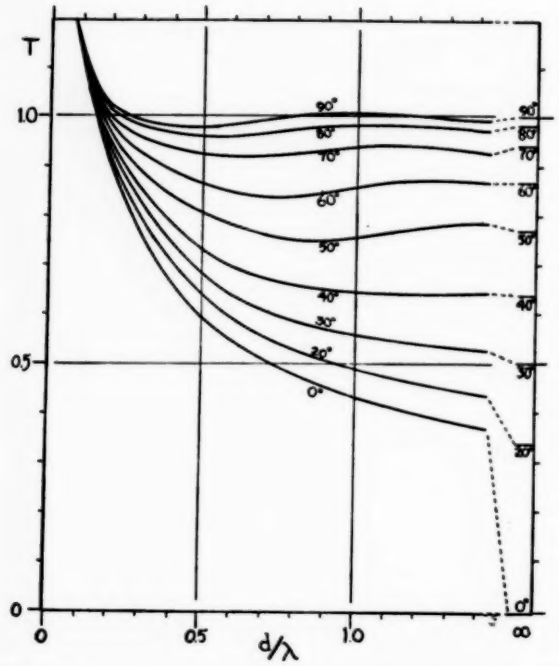

FIG. 4. Corresponding curves for total transmission factor T for the boundary conditions similar to Fig. 3, for different wave-lengths λ and different angles of incidence u.

The total energy scattered by a ribbon in all directions per unit length of ribbon, per unit primary intensity, is obtained by integrating $I'd$ over φ from 0 to 2π, the result being $2dT'$, where

$$T'=(2\pi/c)\sum_m(1/N_m')\sin^2\gamma_m'[So_m(c,\cos u)]^2. \quad (11)$$

The quantity $2T'$ is thus the ratio of the amount of energy actually scattered by the ribbon to the amount of energy which geometrical optics predicts would be reflected by the ribbon at normal incidence. Similarly, for the opposite polarization, T' is the ratio of the energy actually transmitted through the slit to the energy which geometrical optics predicts would be transmitted at normal incidence. The difference of the factor of 2 results from the fact that the wave from the slit covers only the back half of the plane, whereas the wave scattered from the ribbon is on both sides, and at short wave-lengths must be responsible for the reflected wave *and also* for interference with the plane wave behind the ribbon to cause the shadow. This reciprocity between slit and ribbon is an aspect of Babinet's theorem.

The series for I' and T' converge very rapidly for small values of $c=(\pi d/\lambda)$; only the first three or four terms are needed as long as c is less than 4. This range of c covers, however, most of the range not covered by the usual approximate diffraction theory. This range is definitely useful for acoustics and may find some use in the case of short radio waves, although it is not particularly interesting for optics. Curves of I' and T' are given in Figs. 1 and 2. The effects of resonance, for $d=(n\lambda/2)$, are quite noticeable in Fig. 2.

For the complementary boundary conditions of zero value at the surface of a ribbon or zero gradient at the boundary of a slit, the corresponding series for I and T are

$$I=(4\pi/c)\sum_{m,n}(1/N_mN_n)\sin\gamma_m\sin\gamma_n \times\cos(\gamma_n-\gamma_m)Se_m(c,\cos u)Se_n(c,\cos u) \times Se_m(c,\cos\varphi)Se_n(c,\cos\varphi), \quad (12)$$

$$T=(2\pi/c)\sum_m(1/N_m)\sin^2\gamma_m[Se_m(c,\cos u)]^2.$$

These results are useful for calculating the diffraction through a slit of sound waves, or of electromagnetic waves with the magnetic vector parallel to the slit axis; or for calculating the scattering from a ribbon of electromagnetic waves with the electric vector parallel to the ribbon axis. Curves of I and T are given in Figs. 3 and 4. The difference is marked between these curves and those for the other boundary condition; very little resonance effect is noticeable in Fig. 4.

Several other applications of these functions can be made in the field of acoustics. Some of the properties of the ribbon microphone and the ribbon loudspeaker can be studied; and also the absorption of sound by a strip of absorbing material placed in a reflecting wall can be treated. Reports on these applications will be published elsewhere.

The Nature of the Interatomic Forces in Metals

LINUS PAULING
Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California
(Received October 10, 1938)

It has been generally assumed that in the transition elements (Fe, Co, Ni, Cu, etc.) the 3*d* shell is filled with ten electrons or is nearly filled, and that the *d* electrons make no significant contribution to the cohesive forces in metals. Evidence is presented here to show that about half of the *d* orbitals (2.56 of the total of 5) are involved in bond formation, through hybridization with the 4*s* and 4*p* orbitals, and that the number of covalent bonds resonating among the available interatomic positions increases from one to nearly six in the sequence K, Ca, Sc, Ti, V, Cr, remains nearly constant from Cr to Ni, and begins to decrease with Cu. The remaining 2.44 *d* orbitals, with very small interatomic overlapping, are occupied by nonbonding electrons which are mainly responsible for the ferromagnetic and paramagnetic properties of the metals. This point of view provides a qualitative explanation of many properties of the transition metals (including those of the palladium and platinum groups), such as interatomic distance, characteristic temperature, hardness, compressibility, and coefficient of thermal expansion, and it accounts satisfactorily for the observed values of the atomic saturation magnetic moments of the ferromagnetic elements iron, cobalt, and nickel and their alloys. It is also shown to provide a reason for the occurrence of the positive exchange integrals which give rise to ferromagnetism.

DURING the year 1926–1927 I had the great privilege of working as a Guggenheim Fellow in Munich under the direction of Professor Arnold Sommerfeld. At that time he was interested in the systematization of the new quantum mechanics and its application to problems of spectroscopy and atomic structure. It was not until the following year, after the pioneer work of Pauli[1] on the small temperature-independent paramagnetism of the alkali metals, that Professor Sommerfeld and his students began their development of the quantum-mechanical theory of metals.[2]

Their special field of investigation dealt with the electrical and thermal properties of metals. More recently considerable attention has been paid to the question of the nature of the interatomic forces in metals, which are significant for properties such as density, compressibility, crystal energy, and hardness; and it has been found possible to treat this problem in a reasonably satisfactory way for the case of the alkali metals, with a single valence electron per atom.[3] The extension of the theory to other elements has also been attempted.

In recent years I have formed, on the basis mainly of empirical arguments, a conception of the nature of the interatomic forces in metals which has some novel features. According to this view, the metallic bond is very closely related to the ordinary covalent or electron-pair bond; some of the electrons of an atom in a metal are involved with those of neighboring atoms in the interaction described as covalent-bond formation, with the bonds resonating among the available positions in the usual case that the number of positions exceeds the number of bonds. Moreover, all or most of the outer electrons of the atom, *including for the transition elements the d electrons* (3*d*, 4*d*, and 5*d* for the iron, palladium, and platinum transition groups, respectively), take part in bond formation. (It has been assumed by most of the investigators in this field that the *d* electrons make no significant contribution to the cohesive forces.) It is shown below that this point of view leads to a qualitative explanation of many properties of metals, including their magnetic properties.

The combination of two lithium atoms to give the molecule Li_2 is described as involving the formation of a covalent bond between the atoms. In a crystal of fluorine, F_2, the repulsion of the unshared outer electron pairs keeps the molecules spaced so that the minimum intermolecular

[1] W. Pauli, Zeits. f. Physik **41**, 81 (1927).
[2] A. Sommerfeld, Naturwiss. **15**, 825 (1927); A. Sommerfeld, W. V. Houston, and C. Eckart, Zeits. f. Physik **47**, 1 (1928); A. Sommerfeld and N. H. Frank, Rev. Mod. Phys. **3**, 1 (1931); A. Sommerfeld and H. Bethe, *Handbuch der Physik*, Vol. 24, second edition.
[3] J. C. Slater, Phys. Rev. **35**, 504 (1930); **45**, 794 (1934); E. Wigner and F. Seitz, Phys. Rev. **43**, 804 (1933); **46**, 509 (1934); F. Seitz, Phys. Rev. **47**, 400 (1935); K. Fuchs, Proc. Roy. Soc. **A151**, 585 (1935); H. M. Krutter, Phys. Rev. **48**, 664 (1935); M. F. Manning and H. M. Krutter, Phys. Rev. **51**, 761 (1937); W. Shockley, Phys. Rev. **51**, 129 (1937).

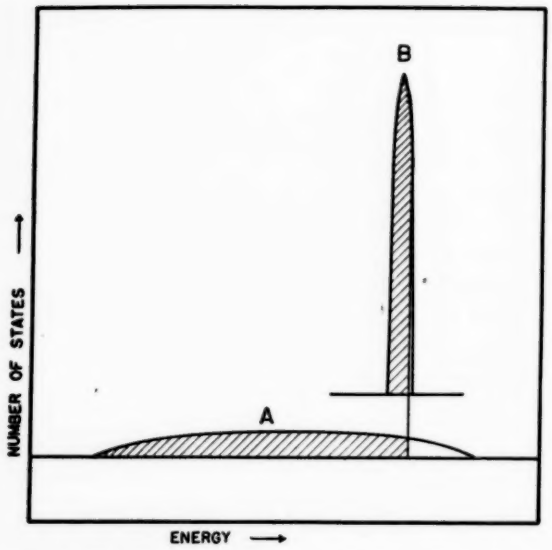

FIG. 1. Electronic states for iron-group atoms, showing number of states as qualitative function of electronic energy. Electrons in band A are paired with similar electrons of neighboring atoms to form bonds. Electrons in band B are d electrons with small interatomic interaction; they remain unpaired until the band is half-filled. The shaded area represents occupancy of the states by electrons in nickel, with 0.6 electron lacking from a completely filled B band. (States corresponding to occupancy of bond orbitals by unshared electron pairs are not shown in the diagram.)

internuclear distance is about twice as great as the intramolecular internuclear distance; there is no chance for resonance of a bond from one pair to another pair of atoms. Moreover, in fluorine all of the orbitals are occupied, so that resonance could occur only by the synchronous interaction of two or more molecules. In a crystal of lithium formed from Li_2 molecules, on the other hand, the conditions for resonance are satisfied; the molecules would be expected to approach one another closely, and each lithium atom has four L orbitals which might be used for bond formation. It is accordingly reasonable to think of the lithium crystal as involving covalent bonds resonating among the available positions. (Around each atom in the body-centered structure there are fourteen positions for bonds, eight at cube corners and six (with internuclear distance 15 percent larger) at cube face centers. On the average one of these is occupied by a bond.) Since the conditions for one-electron bond formation are satisfied,[4] it may be considered that these bonds also make some contribution to the structure of the crystal.

This qualitative description of the interactions in the metal is compatible with quantum mechanical treatments which have been given the problem,[5] and it leads to an understanding of such properties as the ratio of about 1.5 of crystal energy of alkali metals to bond energy of their diatomic molecules (the increase being the contribution of the resonance energy), and the increase in interatomic distance by about 15 percent from the diatomic molecule to the crystal.

The alkaline earth metals, by assuming the configuration $nsnp$, are able to form twice as many bonds as the alkalis. Similarly the succeeding elements in the periodic table can form bonds in increasing number.

The possibility that only some of the outer electrons enter into bond formation needs to be considered. There is a close relation between interatomic distance and bond type, and the values of the interatomic distances in the sequence[6] K, $A2$, 4.62A; Ca, $A1$, $A3$, 3.93–3.95A; Sc, not yet investigated; Ti, $A3$, 2.92–2.95A; V, $A2$, 2.63A; Cr, $A2$, 2.49A, as well as those in the similar sequences from Rb to Mo and Cs to W, indicate that the number of bonds resonating among the available positions increases from one to nearly six.

This conclusion is given strong support by the magnetic properties of the metals. Let us consider vanadium as an example. The configuration of the normal vanadium atom is $3d^3 4s^2$. If only the two $4s$ electrons were involved in bond formation (perhaps with promotion of one to $4p$) the core of the atom, with the configuration $3d^3$, would have a large magnetic dipole moment,[7] and in consequence the metal would show ferromagnetism or large paramagnetism, decreasing in magnitude with increasing temperature. Instead vanadium is only weakly paramagnetic (molal susceptibility at room temperature 60×10^{-6} cgsmu.) and its paramagnetism is nearly temperature-independent; there is, indeed, a small in-

[4] L. Pauling, J. Am. Chem. Soc. **53**, 3225 (1931).

[5] See in particular J. C. Slater, Phys. Rev. **35**, 509 (1930).

[6] The symbols $A1$, $A2$, and $A3$ represent the three simple metal structures: cubic closest packed, body centered, and hexagonal closest packed, respectively.

[7] Values 3.8 to 4.0 Bohr magnetons are observed for iron-group ions with this configuration.

crease in susceptibility with increasing temperature. This behavior is that expected if all five outer electrons take part in bond formation, with pairing of spins as in the usual electron-pair bond. The observed paramagnetic properties are then explained, as in Pauli's theory of the temperature-independent paramagnetism of the alkali metals, by the assumption that some of the electrons occupy excited levels with unpaired spins (that is, that under equilibrium conditions some of the bonds are broken), the number of these unpaired electrons increasing with increasing temperature. Since diamagnetism or small temperature-independent paramagnetism is observed for all of the metals K, Ca, Ti, V, Cr, Rb, Sr, Zr, Cb, Mo, Cs, Ba, La, Ta, and W, it seems necessary to accept the conclusion that all of the outer electrons take part in bond formation.

The *d* orbitals alone are not especially well suited to use in bond formation, but hybridization of *d*, *s*, and *p* orbitals leads to the best bond orbitals known,[8] such as the d^2sp^3 orbitals in octahedral complexes ($[Fe(CN)_6]^{=}$, $[PtCl_6]^{=}$, etc.). There are available a total of nine relatively stable orbitals: five 3*d*, one 4*s*, and three 4*p* for the iron-group elements, and corresponding sets for the other series; and there exists accordingly the possibility that as many as nine outer electrons per atom enter into bond formation. The number of bonds would then be six for chromium, seven for manganese, eight for iron, and nine for cobalt; at nickel, with ten electrons, one unshared pair would occupy one of the nine orbitals, leaving only eight bonds, and the continued decrease in the number of bonds would lead to seven for copper, six for zinc, five for gallium, four for germanium, three for arsenic, two for selenium, and one for bromine.

It is indicated by the observed interatomic distances and shown by magnetic data that there occurs some deviation from this simple and attractive scheme in the middle region of the sequence. From chromium to cobalt the interatomic distances do not continue to decrease in value, as expected with increase in the number of bonds; instead they remain nearly constant: Cr, *A*2, 2.49A; Mn, no simple structure; Fe, *A*2, 2.48A, *A*1, 2.52A; Co, *A*1, *A*3, 2.50–2.51A; Ni, *A*1, *A*3, 2.49A; Cu, *A*1, 2.55A; Zn, *A*3, 2.66, 2.91A. This indicates that there are only about six bonds per atom for iron, cobalt, and nickel, and still fewer for copper and the succeeding elements.

The values for the atomic saturation magnetization at the absolute zero, σ_A, for the ferromagnetic metals iron, cobalt, and nickel are 2.22, 1.71, and 0.61 Bohr magnetons per atom, respectively.[9] These numbers are the average numbers of unpaired electron spins in the metals (the approximation of the *g* factor to 2 found in gyromagnetic experiments shows that the orbital moment is nearly completely quenched, as in complex ions containing the transition elements).

In the past the following interpretation has been given these numbers.[10] In nickel, for example, it has been assumed that the 3*d* shell is occupied by 9.39 (on the average) of the total of ten outer electrons, with 0.61 electron in the 4*s* shell. The 0.61 *s* electron alone was assumed to interact strongly between atoms, and to give rise to the principal cohesive forces in the metal. This interaction would quench the magnetic moments of the *s* electrons almost completely. The hole of 0.61 electron in the nearly completed *d* shell was then assumed to give a corresponding magnetic moment to the atom, the interaction of *d* orbitals of adjacent atoms being considered to be small.[11] It was similarly assumed that the cohesion of cobalt is due to 0.71 *s* electron, that

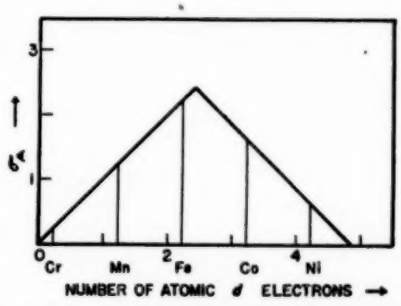

FIG. 2. Predicted dependence of atomic saturation magnetic moment σ_A as function of number of atomic *d* electrons.

[8] L. Pauling, J. Am. Chem. Soc. **53**, 1367 (1931).

[9] P. Weiss and R. Forrer, Ann. de physique **12**, 279 (1929); E. C. Stoner, *Magnetism and Matter* (Methuen and Co., London, 1934), p. 366.

[10] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Oxford University Press, 1936), pp. 190, 222, 316.

[11] J. C. Slater, Rev. Mod. Phys. **6**, 272 (1934), has also concluded that the *d* electrons do not take much part in cohesion.

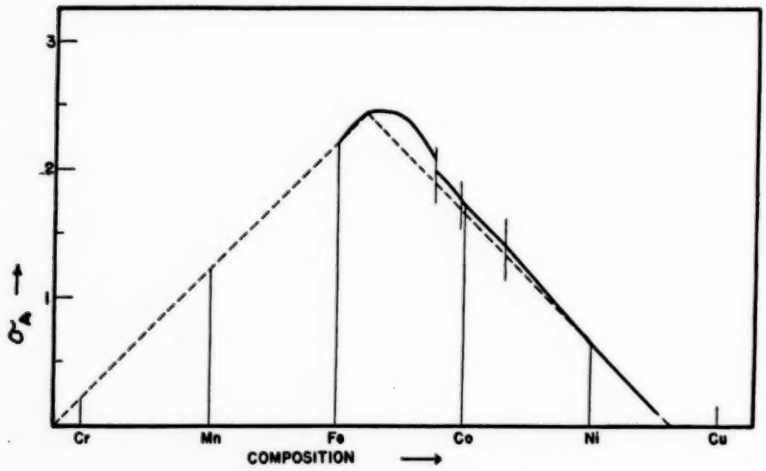

FIG. 3. Comparison of experimental values (solid curves) and predicted values (dashed lines) of σ_A for Fe-Co, Co-Ni, and Ni-Cu alloys. The short vertical lines indicate change in crystal structure.

of iron to 0.22 *s* electron, and that of copper to one *s* electron (the *d* shell for copper having its full complement of ten electrons). The difficulties of accounting for the interatomic distances and mechanical properties of the metals in this way are obvious.

Our picture, on the other hand, is that some of the 3*d* orbitals, as well as the 4*s* and 4*p* orbitals, are used for the formation of five or six bonds per atom. Instead of assuming the configuration $3d^{10}4s$ for copper, with bonding by the 4*s* electron only, we assume a configuration such as $d^2d^2d^2ddspp$ (to list the independent orbitals separately), so that each atom contains three unshared pairs in the *d* shell and five electrons used in forming bonds with surrounding atoms. This structure, involving only unshared and shared electron pairs, accounts satisfactorily for the observed diamagnetism of the metal.

Our knowledge of the properties of orbitals indicates that some of the 3*d* orbitals might be combined with the 4*s* and 4*p* orbitals to form bond orbitals in metals, the other 3*d* orbitals being unsuited to bond formation, but does not suffice to give a theoretical derivation of the number of *d* orbitals in each of these classes. Empirical evidence, outlined below, indicates that about 2.44 *d* orbitals (on the average) show only weak interatomic interactions, and that the remaining 2.56 *d* orbitals combine with the *s* orbital and the *p* orbitals to form hybrid bond orbitals.

In Fig. 1 there is indicated the division of the nine outer orbitals into these two classes. It is assumed that electrons occupying orbitals of the first class (weak interatomic interactions) in an atom tend to remain unpaired (Hund's rule of maximum multiplicity), and that electrons occupying orbitals of the second class pair with similar electrons of adjacent atoms. Let us call these orbitals *atomic orbitals* and *bond orbitals*, respectively. In copper all of the atomic orbitals are occupied by pairs. In nickel, with $\sigma_A = 0.61$, there are 0.61 unpaired electrons in atomic orbitals, and in cobalt 1.71. (The deviation from unity of the difference between the values for cobalt and nickel may be the result of experimental error in the cobalt value, which is uncertain because of the magnetic hardness of this element.) This indicates that the energy diagram of Fig. 1 does not change very much from metal to metal. Substantiation of this is provided by the values of σ_A for copper-nickel alloys,[12] which decrease linearly with mole fraction of copper from $\sigma_A = 0.61$ for pure nickel toward the value zero for mole fraction 0.6 of copper, and by the related values for zinc-nickel and other alloys.[13] The value $\sigma_A = 2.61$ would accordingly be expected for iron, if there were 2.61 or more *d* orbitals in the atomic orbital class. We conclude from the observed value $\sigma_A = 2.22$ for iron that

[12] M. Alder, *Dissertation* (Zurich, 1916), quoted by Stoner, reference 9, p. 537.

[13] C. Sadron, Ann. de physique **17**, 371 (1932).

the number of orbitals in this class is less than 2.61, and lies between 2.22 and 2.61.

If it be assumed that one electron is removed from these atomic orbitals at each step Ni-Co-Fe, the number of orbitals is determined as 2.44 ± 0.03. There is shown as Fig. 2 a diagram representing the expected change of atomic saturation moment σ_A with change in number of electrons. It is seen that σ_A is predicted to rise to a maximum value of 2.44 Bohr magnetons at a point 23 percent of the way from iron to cobalt. It is indeed observed[14] that the atomic saturation moment for iron-cobalt alloys rises to its maximum value, 2.47 Bohr magnetons, at 26 atomic percent cobalt. Curves showing measured values of σ_A for iron-cobalt, cobalt-nickel, and nickel-copper alloys are compared with the theoretical curve for 2.44 atomic d orbitals in Fig. 3. The experimental values are seen to indicate a tendency for the class of atomic d orbitals to remain half-filled with electrons (maximum multiplicity) over a small range of composition; the values of σ_A remain close to the maximum until an excess of about 0.25 electron is present, and then slowly decrease to the expected line.

There are no data available for testing the predicted numbers 0.22 and 1.22 of unpaired electron spins for chromium and manganese, respectively.

It seems probable that the palladium and platinum transition groups approximate the iron group very closely. For example, the paramagnetic susceptibility of palladium-gold alloys falls to zero at 60 atomic percent gold, the paramagnetic susceptibility of palladium containing hydrogen becomes zero at 60 atomic percent hydrogen, and the saturation moment of nickel is changed very little by the addition of palladium.[15]

It is of interest to calculate the number of bonding electrons, which we assume for the elements Cr to Ni to be all of the outer electrons except those in the atomic d orbitals. For chromium this number is $6-0.22=5.78$, and the same value is found for the other atoms also; this supports the evidence from interatomic distances that there is little change in bond type from chromium to nickel.

There are 2.56 d orbitals available for bond formation. To form 5.78 bonds these would hybridize with the s orbital and 2.22 of the less stable p orbitals. In copper, with one electron more than nickel, there is available an additional 0.39 electron after the hole in the atomic d orbitals is filled. This might take part in bond formation, with use of additional $4p$ orbital. However, the increase in interatomic distance from nickel to copper suggests that it forms part of an unshared pair with part of the bonding electrons, thus decreasing the effective number of bonds.

To this treatment of the metals the objection might be raised that it requires the considerable promotion of electrons to higher levels (for the separated atoms): thus, whereas Mott and Jones[16] consider that only the configurations $3d^8 4s^2$, $3d^9 4s$, and $3d^{10}$, which give states lying within an energy range of about 1.5 ev, can be considered for nickel, our treatment requires consideration of configurations such as $d^7 sp^2$, which probably lie about 10 ev above the normal state. It is not the energy of the configuration of the isolated atom which is significant, however, but that of the metal itself, and the additional bond energy of the metal with this configuration compensates for the instability of the atom. The crystal energy (heat of sublimation) of a metal provides little information regarding the structure of the metal because it is the difference of two quantities, the energy of the metal and the energy of the separated atoms. It is properties such as interatomic distance, compressibility, and coefficient of thermal expansion, characteristic of the metal itself, which give information about the nature of the bonds.

We have mentioned that the increase in effective number of bonds from one to about six in the series of metals K to Cr, its approximate constancy to Ni, and subsequent decrease are compatible with the observed course of inter-

[14] P. Weiss and R. Forrer, Ann. de physique **12**, 279 (1929). Mention should also be made of the moments found for iron, cobalt, and nickel by measurement of the paramagnetic susceptibility above the Curie point, the values being 3.15–3.20, 3.15, and 1.61–1.78 Bohr magnetons, respectively (W. Sucksmith and R. R. Pearce, Proc. Roy. Soc. **A167**, 189 (1938)). The root-mean-square values predicted from the magneton numbers on the assumption that the iron atoms contain either two or three unpaired electron spins, the cobalt atoms one or two, and the nickel atoms zero or one are 3.09, 2.56, and 1.35 Bohr magnetons, respectively. The reason for the only very approximate agreement is not known.

[15] Mott and Jones, reference 10, pp. 197, 199, 200.

[16] Reference 10, p. 190.

atomic distances for this sequence of elements. A similar behavior is observed for the sequences Rb to Ag and Cs to Au. In each of these three sequences the maximum or minimum values (corresponding to maximum strength of bonds) of the compressibility, coefficient of thermal expansion, characteristic temperature, melting point, and hardness occur for the metals to which we have assigned the maximum number of bonds.[17]

In the development of the quantum mechanical theory of ferromagnetism by Heisenberg, Frenkel, Bloch, Slater, and other investigators the assumption has been made that the exchange integrals between the unpaired *d* electrons of adjacent atoms are positive in sign. It has in the past been very difficult to understand why this should occur, since exchange integrals between atoms are usually negative. The situation is clarified by our picture of the structure of the transition metals. A large negative value of an exchange integral (corresponding to a strong covalent bond) results from the large overlapping of the orbitals of two atoms. Good bond orbitals can be formed by hybridization of the outer *d*, *s*, and *p* orbitals of the metal atoms, and we expect that all the *d* orbitals which are able to form strong bonds will enter into bond formation. The remaining *d* orbitals will hence be those which avoid overlapping with the orbitals of adjacent atoms. The exchange integrals of these atomic *d* orbitals will accordingly be small in absolute value, and may be positive.[18] This conclusion agrees with the observation that the transition metals may be either paramagnetic or ferromagnetic, the absolute magnitude of the exchange integrals in either case being about 1 kcal./mole.

[17] The minimum reported values of the compressibility occur for the three sequences at Ni, Mo, and Pt, respectively; the minima of the coefficient of thermal expansion at Fe, Pd, and W; the maxima of the characteristic temperature at Cr, Ru, and W or Ir; those of the melting point at Ti, Mo, and W; and those of the hardness at Cr, Ru, and Os.

[18] Examination of the form of the exchange integral shows that it is necessarily positive if the two orbitals are orthogonal.

It is obvious that many factors may enter into the determination of the sign of the exchange integral.[19] We would predict that in an atom which formed many strong bonds with its neighbors the *d* orbitals with negative exchange integrals would be drawn into bond formation, and that unpaired electrons in the remaining *d* orbitals could give rise to ferromagnetism. The condition for ferromagnetism that the metal atoms form a large number of strong bonds is satisfied by the ferromagnetic elements. We would also expect that manganese would develop ferromagnetic properties in alloys in which its atoms form more and stronger bonds with surrounding atoms than in the element itself. This may occur in the Heusler alloys (such as Cu_2AlMn and Cu_2SnMn).

In this discussion of the transition elements we have considered only the orbitals $(n-1)d\ ns\ np$. It seems probable that in some metals use is made also of the *nd* orbitals in bond formation. In gray tin, with the diamond structure, the four orbitals $5s5p^3$ are used with four outer electrons in the formation of tetrahedral bonds, the $4d$ shell being filled with ten electrons. The structure of white tin, in which each atom has six nearest neighbors (four at 3.016A and two at 3.175A), becomes reasonable if it is assumed that one of the $4d$ electrons is promoted to the $5d$ shell, and that six bonds are formed with use of the orbitals $4d5s5p^35d$.

This arrangement of bonds for white tin is not to be considered as the only one contributing to the structure of the metal. An atom in a metal shows great versatility in the formation of bonds. There is little difference in stability of structures involving not only different numbers of bonds but also varied types of coordination, and in consequence an atom in a metal can form bonds after deformation of the original structure which are approximately as strong as the original bonds. This gives a metal its ability to heal itself after deformation, which underlies such characteristic metallic properties as malleability and ductility.

[19] This question has been discussed by J. C. Slater, Phys. Rev. **36**, 57 (1930).

Elastic Scattering of Yukawa Particles. I

OTTO LAPORTE
University of Michigan, Ann Arbor, Michigan
Dedicated to A. Sommerfeld Upon His Seventieth Birthday
(Received October 14, 1938)

Elastic scattering of the heavy particle obeying Bose statistics is examined by the relativistically invariant equations recently developed by Proca and by Dirac. Perturbation calculus of the "variation of constants" type is applied. For the evaluation of certain sums occurring in the scattering cross sections, destruction operators are constructed similar to those first developed by Weisskopf for the Dirac equation. While according to Mott's calculation of the scattering of electrons a polarization effect does not occur until the second approximation, such an effect appears in our case even in the first approximation. The second approximation will be treated in a later paper.

I

WITHIN the last year a heavy particle possessing spin unity and obeying Bose statistics has been introduced by various authors[1] with considerable success for the explanation of nuclear forces. Originally this particle was thought to be of the Pauli-Weisskopf type possessing spin zero,[2] but the resulting nuclear forces proved to be of the wrong sign.[3] Thus for the first time these elementary particles find application whose wave equations belong to the family recently developed by Dirac,[4] for the case of $s>\frac{1}{2}$. While the comparative simplicity of the wave equations of such particles seems to demolish the unique position of the wave equation of particles with $s=\frac{1}{2}$, the closer analysis of these new equations reveals that certain simplifying facts—e.g. the possibility of the construction of only one stress-energy tensor[5]—are common only to particles of spin 0, $\frac{1}{2}$ and 1. Thus one is perhaps justified in conjecturing that only these latter particles will find permanent places in the theory. In this connection it is of interest to note that the wave equations for particles with spin 1 have been found and discussed independently although from a different point of view, by Proca.[6]

The various effects occurring when such particles interact with matter, for instance with protons, are bremstrahlung, inelastic scattering through exchange, and purely elastic Rutherfordian scattering. The first two effects have been discussed by the authors mentioned under reference 1. The third effect will be treated in this paper. Although we regard it as improbable that this elastic scattering will ever be separated from the other two effects, it seems to us necessary to understand quantitatively all effects through which a heavy particle can interact with matter. Furthermore, since spin terms are of considerable influence upon the result, it may perhaps be possible in the future to decide between the two alternatives of $s=0$ or $s=1$ by performing a scattering experiment.

The corresponding computation for electrons was carried out in considerable detail by Mott[7] and was later very much simplified by Sauter.[8] In this latter paper a Born approximation was used and the calculations were carried out leaving the scalar force field through which the impinging particle and the nucleus interact, completely general. By not specializing this general field to Coulomb interaction, until the final result, it was possible to avoid all convergence difficulties. In this way the first Born approximation gives—

[1] Yukawa, Proc. Phys. Math. Soc. Japan **17**, 48 (1935); Yukawa and Sakata, Proc. Phys. Math. Soc. Japan **19**, 1084 (1937); Yukawa, Sakata and Taketani, Proc. Phys. Math. Soc. Japan **20**, 319 (1938); Kemmer, Proc. Roy. Soc. **A166**, 127 (1938); Fröhlich, Heitler and Kemmer, Proc. Roy. Soc. **A166**, 154 (1938); Bhabha, Proc. Roy. Soc. **A166**, 501 (1938); Heitler, Proc. Roy. Soc., **A166**, 529 (1938).
[2] Pauli and Weisskopf, Helv. Phys. Acta **7**, 709 (1934).
[3] Third and fourth paper of reference 1.
[4] Dirac, Proc. Roy. Soc. **A155**, 447 (1936).
[5] Fierz, Helv. Phys. Acta 1938, in press.
[6] Proca, J. de phys. et rad. **7**, 347 (1936).
[7] N. F. Mott, Proc. Roy. Soc. **A124**, 425 (1929); **135**, 429 (1932).
[8] F. Sauter, Ann. d. Physik **18**, 61 (1933).

except for relativistic corrections—the Rutherford term, while polarization terms appear but in the second approximation.

In the following sections the same treatment is applied to the Dirac-Proca equations for particles of spin unity. In Section II the unperturbed wave functions are discussed and various preparatory computations are carried out. In order to exploit fully the orthogonality properties of the eigenfunctions, we shall find it necessary to construct operators analogous to the destruction or annihilation operators which Weisskopf[9] and others have set up for the Dirac equations. In Section III perturbation theory of the variation-of-constants type is developed and expressions for the cross section of scattered particles are obtained in first and second approximation. In Section IV the first approximation is evaluated and discussed. A report on the evaluation of the second approximation will be given in a later paper.

II

(a) In Yukawa's notation the Dirac-Proca equations are

$$\left(\frac{1}{c}\frac{\partial}{\partial t}+\frac{i}{\hbar c}V\right)\mathbf{F}-\text{curl }\mathbf{G}-K\mathbf{U}=0,$$
$$\left(\frac{1}{c}\frac{\partial}{\partial t}+\frac{i}{\hbar c}V\right)\mathbf{U}+\text{grad }U_0+K\mathbf{F}=0, \tag{1}$$

$$\text{div }\mathbf{F}+KU_0=0,$$
$$\text{curl }\mathbf{U}-K\mathbf{F}=0. \tag{2}$$

$V(x, y, z)$ is the potential energy through which the (stationary) nucleus interacts with the arriving particle. $K=Mc/\hbar$ is its reciprocal Compton wave-length. The wave functions $\mathbf{F}$, $\mathbf{G}$, $\mathbf{U}$ are complex vectors, while U_0 is a complex scalar. The wave equations (1), (2) are relativistically invariant, but as we shall not make use of this property it does not seem necessary to introduce a four-dimensional index notation. Suffice it to say that actually $\mathbf{F}$ and $\mathbf{G}$ form an antisymmetric tensor, and $\mathbf{U}$ and U_0 a four-vector.

In order to get rid of Eqs. (2) which play the role of accessory conditions, we eliminate $\mathbf{G}$ and U_0 with the result:

$$K\left(\frac{1}{c}\frac{\partial}{\partial t}+\frac{i}{\hbar c}V\right)\mathbf{F}-\text{grad div }\mathbf{U}$$
$$+\Delta\mathbf{U}-K^2\mathbf{U}=0, \tag{3}$$

$$K\left(\frac{1}{c}\frac{\partial}{\partial t}+\frac{i}{\hbar c}V\right)\mathbf{U}-\text{grad div }\mathbf{F}\quad+K^2\mathbf{F}=0. \tag{4}$$

In this form each equation contains a first derivative with respect to the time which is just the form required by transformation theory. We are now ready to apply the method of the variation of constants, regarding V as perturbation parameter.

An alternative approach is to eliminate $\dot{\mathbf{F}}$ in Eq. (3) by means of Eq. (4), then to assume a definite stationary time dependence with an energy E. Then equations of the following type result:

$$\{\Delta+(E/\hbar c)^2-K^2\}\mathbf{U}=\text{linear vector functions of } \mathbf{U} \text{ and of } \mathbf{F} \text{ whose coefficients involve up to second derivatives of } V. \tag{5}$$

Similarly for $\mathbf{F}$. These Poisson-type equations may then be solved by successive approximations in the familiar way by means of Green functions.[10] Because of the complicated structure of the right side of the above equation it was found to be simpler to work with (3) and (4).

(b) We now have to study the unperturbed system in some detail. The unperturbed wave vectors i.e., the solutions of (3) and (4) for $V=0$ we shall write ${}_0\mathbf{U}$ and ${}_0\mathbf{F}$. We then enter these equations with the following plane wave expressions

$${}_0\mathbf{U}=(2\pi)^{-\frac{3}{2}}\mathbf{u}\exp\left[i(\mathbf{k}\cdot\mathbf{x})-iEt/h\right],$$
$${}_0\mathbf{F}=(2\pi)^{-\frac{3}{2}}\mathbf{f}\exp\left[i(\mathbf{k}\cdot\mathbf{x})-iEt/h\right]. \tag{6}$$

$\mathbf{u}$ and $\mathbf{f}$ are as yet unknown vectorial wave amplitudes which may yet depend upon $\mathbf{k}$, the wave vector, and upon E, the energy. The secular determinant for E can be calculated:

$$\left[(E/\hbar c)^2-K^2-k^2\right]^3=0. \tag{7}$$

It has two triple roots $E/\hbar c=\pm(K^2+k^2)^{\frac{1}{2}}$ which

[9] Weisskopf, Zeits. f. Physik **89**, 27 (1934).

[10] This is the original method of Born and Wentzel. See Sommerfeld, *Wave Mechanics*, p. 192, also *Handbuch der Physik*, Vol. 24_1, p. 712–721.

correspond to the states of positive and negative energy. We shall distinguish between these in the following manner: Let

$$\epsilon = +(K^2+k^2)^{\frac{1}{2}} \tag{8}$$

be the positive value of the energy divided by $\hbar c$. Let η be a quantum number which is capable of only the values $+1$ and -1. We shall then write the argument of the exponential function in (6)

$$i(\mathbf{k}\cdot\mathbf{x}) - ic\eta\epsilon t \quad \text{with} \quad \eta = \pm 1. \tag{9}$$

The fact that each root of (7) is threefold, can be described, as has been pointed out by various authors, by ascribing a longitudinal and two transverse normal vibrations to both ${}_0\mathbf{U}$ and ${}_0\mathbf{F}$. In accordance with this fact it is convenient to introduce a system of mutually perpendicular unit vectors $\mathbf{e}(p, \mathbf{k})$ of the following kind:[11]

$$\mathbf{e}(0, \mathbf{k}) = \mathbf{k}/|k|, \tag{10}$$

$$(\mathbf{e}^*(p, \mathbf{k})\cdot\mathbf{e}(p', \mathbf{k})) = \delta_{pp'}. \tag{11}$$

The unit vectors $\mathbf{e}(\mathrm{I})$, $\mathbf{e}(-\mathrm{I})$ are evidently complex, in particular

$$\mathbf{e}^*(\mathrm{I}) = \mathbf{e}(-\mathrm{I}).$$

They correspond to right or left circular polarization. One can also use a system of real unit vectors, namely

$$\begin{aligned} \mathbf{e}(1) &= 2^{-\frac{1}{2}}(\mathbf{e}(\mathrm{I}) + \mathbf{e}(-\mathrm{I})), \\ \mathbf{e}(-1) &= 2^{-\frac{1}{2}}i^{-1}(\mathbf{e}(\mathrm{I}) - \mathbf{e}(-\mathrm{I})), \end{aligned} \tag{12}$$

which correspond to linear polarization. We shall for the present use the complex system.

Utilizing the notation defined by Eqs. (8) to (11) we may now refine our Ansatz (6), to:

$$\begin{aligned} {}_0\mathbf{U}(\eta, p, \mathbf{k}; \mathbf{x}, t) &= (2\pi)^{-\frac{1}{2}}\mathbf{u}(\eta, p, \mathbf{k}) \\ &\quad \times \exp\left[i(\mathbf{k}\cdot\mathbf{x}) - ic\eta\epsilon t\right], \\ {}_0\mathbf{F}(\eta, p, \mathbf{k}; \mathbf{x}, t) &= (2\pi)^{-\frac{1}{2}}f(\eta, p, \mathbf{k}) \\ &\quad \times \exp\left[i(\mathbf{k}\cdot\mathbf{x}) - ic\eta\epsilon t\right], \end{aligned} \tag{13}$$

with $\eta = \pm 1$; $p = \pm\mathrm{I}$ or 0. The amplitude vectors $\mathbf{u}$ and $\mathbf{f}$ have yet to be determined. Entering (3) and (4) for the case of $V = 0$ and using (8) we have:

$$\begin{aligned} -i\eta\epsilon K\mathbf{f}(\eta p\mathbf{k}) + \mathbf{k}(\mathbf{k}\cdot\mathbf{u}(\eta p\mathbf{k})) \\ \qquad - \epsilon^2\mathbf{u}(\eta p\mathbf{k}) = 0, \\ -i\eta\epsilon K\mathbf{u}(\eta p\mathbf{k}) + \mathbf{k}(\mathbf{k}\cdot\mathbf{f}(\eta p\mathbf{k})) \\ \qquad + K^2\mathbf{f}(\eta p\mathbf{k}) = 0. \end{aligned} \tag{14}$$

(Round brackets and dot signify the scalar product.) This system decomposes into three pairs of equations as one sees upon putting

$$\begin{aligned} \mathbf{u}(\eta, p, \mathbf{k}) &= u(\eta, p, k)\mathbf{e}(p, \mathbf{k}), \\ \mathbf{f}(\eta, p, \mathbf{k}) &= f(\eta, p, k)\mathbf{e}(p, \mathbf{k}) \end{aligned} \tag{15}$$

and considering (10) and (11). We will write the solutions

$$\begin{aligned} u(\eta pk) &= K\epsilon\Delta(\eta pk), \\ f(\eta pk) &= i\eta(K^2+p^2k^2)\Delta(\eta pk) \end{aligned} \tag{16}$$

where $\Delta(\eta pk)$ is an as yet arbitrary multiplicative constant.

(c) We now have to study the orthogonality properties of the unperturbed solutions (13), (15), (16) and obtain a convenient normalization which will determine $\Delta(\eta pk)$. For this purpose it is necessary to derive the continuity equation. Multiply Eq. (3) scalarly with $\mathbf{U}^*$, then Eq. (4) with $-\mathbf{F}^*$ and add. Then subtract from this expression its complex conjugate. The result can be written

$$c^{-1}\dot{\rho} + \mathrm{div}\,\mathbf{j} = 0, \quad \rho = i^{-1}\{(\mathbf{U}^*\cdot\mathbf{F}) - (\mathbf{F}^*\cdot\mathbf{U})\}. \tag{17}$$

The expression for $\mathbf{j}$ we shall not write down, as we are not going to make any use of it. It is important to be aware of the fact that this is really the continuity equation and not perhaps the time component of the equation expressing the conservation of energy and momentum.

Equation (17) furnishes in the usual fashion an orthogonality condition. When substituting in the expression for ρ the undisturbed eigenfunctions, *viz.* ${}_0\mathbf{U}^*(\eta, p, \mathbf{k}; \mathbf{x}, t)$ and ${}_0\mathbf{F}^*(\eta, p, \mathbf{k}; \mathbf{x}, t)$ for $\mathbf{U}^*$ and $\mathbf{F}^*$, and ${}_0\mathbf{U}(\eta', p', \mathbf{k}'; \mathbf{x}, t)$ and ${}_0\mathbf{F}(\eta', p', \mathbf{k}'; \mathbf{x}, t)$ for $\mathbf{U}$ and $\mathbf{F}$, integration over $\mathbf{x}$ space gives the δ function $\delta(\mathbf{k}'-\mathbf{k})$. For the remaining expression, i.e., the ρ symbol formed with the vector amplitudes (15), we introduce because of its frequent reappearance the sym-

[11] See H. A. Kramers, *Quantentheorie des Elektrons und der Strahlung*, Vol. 2, p. 259.

bol $\bar{\rho}$:

$$(\eta' p' \mathbf{k}' | \bar{\rho} | \eta p \mathbf{k}) = i^{-1}(u_\alpha^*(\eta' p' \mathbf{k}') f_\alpha(\eta p \mathbf{k}) - f_\alpha^*(\eta' p' \mathbf{k}') u_\alpha(\eta p \mathbf{k})). \quad (18)$$

The Greek indices signify (cartesian) vector (or tensor) components; doubly occurring Greek indices are to be summed from 1 to 3. For $\mathbf{k}=\mathbf{k}'$ the above quantity vanishes because of the orthogonality of the $\mathbf{e}(p, \mathbf{k})$ vectors (Eq. (11)) except if $p=p'$. In this case, one has, using (16):

$$(\eta', p, \mathbf{k} | \bar{\rho} | \eta, p, \mathbf{k}) = \epsilon K(K^2+p^2K^2) \times (\eta'+\eta)\Delta^*(\eta', p, k)\Delta(\eta p k),$$

which vanishes unless $\eta'=\eta$. We can use this fact for the determination of $\Delta(\eta p k)$, by normalizing $(\eta p \mathbf{k} | \bar{\rho} | \eta p \mathbf{k})$ to have the value η. This leads, except for an arbitrary phase factor, to the following expression:

$$\Delta(\eta p k) = [2\epsilon K(K^2+p^2k^2)]^{-\frac{1}{2}}. \quad (19)$$

We summarize the above consideration by the orthogonality relation:

$$(\eta' p' \mathbf{k}' | \bar{\rho} | \eta p \mathbf{k}) = \eta \delta_{\eta\eta'} \delta_{pp'}. \quad (20)$$

The fact that the charge is not positive definite is completely analogous to the case of $s=0$ i.e., of the scalar Gordon-Klein-Schrödinger equation. For this latter equation this was first pointed out by Pauli and Weisskopf.[2]

The converse orthogonality relations are obtained in the following manner: Multiply (20) by $\eta' f_\beta(\eta' p' \mathbf{k})$ and sum over η' and p'. β indicates any cartesian component. The following relations result:

$$\sum_{\eta, p} \eta f_\alpha^*(\eta p \mathbf{k}) f_\beta(\eta p \mathbf{k}) = 0,$$

$$\sum_{\eta, p} \eta u_\alpha^*(\eta p \mathbf{k}) u_\beta(\eta p \mathbf{k}) = 0, \quad (21)$$

$$\sum_{\eta, p} \eta u_\alpha^*(\eta p \mathbf{k}) f_\beta(\eta p \mathbf{k}) = i\delta_{\alpha\beta}.$$

(d) The completeness relation may be easily derived. A general solution of the (unperturbed) wave equations (3) and (4) is:

$$\mathbf{U}(\mathbf{x}, t) = \sum_{\eta, p} \int d\mathbf{k} A(\eta, p, \mathbf{k})\, {}_0\mathbf{U}(\eta, p, \mathbf{k}; \mathbf{x}, t),$$

$$\mathbf{F}(\mathbf{x}, t) = \sum_{\eta, p} \int d\mathbf{k} A(\eta, p, \mathbf{k})\, {}_0\mathbf{F}(\eta, p, \mathbf{k}; \mathbf{x}, t). \quad (22)$$

$d\mathbf{k}$ stands for $dk_1 dk_2 dk_3$, $\mathbf{x}$ for (x, y, z). We compute the Integral

$$\int d\mathbf{x} i^{-1}\{(\mathbf{U}^* \cdot \mathbf{F}) - (\mathbf{F}^* \cdot \mathbf{U})\}.$$

Introducing into it (22) and applying (1) $\mathbf{k}$ orthogonality, (2) p orthogonality, and (3) η orthogonality according to (20) we obtain

$$\int d\mathbf{x} i^{-1}\{(\mathbf{U}^* \cdot \mathbf{F}) - (\mathbf{F}^* \cdot \mathbf{U})\} = \int d\mathbf{k} \sum_{\eta, p} \eta |A(\eta, p, \mathbf{k})|^2. \quad (23)$$

(e) In later sections we shall make considerable use of so-called "destruction operators" or "annihilation operators" analogous to those introduced by Weisskopf[12] into the theory of the Dirac electron. There are two kinds of destruction operators. The first kind has the eigenvalue $+1$ for a state with a definite polarization quantum number p and zero for states with any other polarization quantum number. The second kind has the eigenvalue $+1$ for a state of positive energy (i.e., with $\eta=+1$) and zero for the state with negative energy.

We shall first deal with the former kind. Omitting for the moment the quantum number η which is of no importance, we have because of (15):

$$\frac{i}{|k|}[\mathbf{k} \times \mathbf{u}(p)] = p\mathbf{u}(p); \quad \text{eigenvalues } +1, 0, -1.$$

$$-\frac{1}{k^2}[\mathbf{k} \times [\mathbf{k} \times \mathbf{u}(p)]] = p^2\mathbf{u}(p); \quad \text{eigenvalues } +1, 0, +1.$$

We might have written $\mathbf{f}(p)$ as well; it is immaterial upon which amplitude vector we are operating. With these two operators and the identity operator one can construct three operators for which the relation holds:

$$D_{\alpha\beta}(p_0, \mathbf{k}) u_\beta(p, k) = \delta_{p, p_0} u_\alpha(p, \mathbf{k}). \quad (24)$$

These are:

$$D_{\alpha\beta}(\pm \mathbf{I}, \mathbf{k}) = \pm \frac{1}{2i|k|} \epsilon_{\alpha\beta\gamma} k_\gamma + \tfrac{1}{2}\delta_{\alpha\beta} - \frac{1}{2k^2} k_\alpha k_\beta, \quad (25)$$

$$D_{\alpha\beta}(0, \mathbf{k}) = \frac{1}{k^2} \mathbf{k}_\alpha \mathbf{k}_\beta.$$

[12] See reference 9.

$\epsilon_{\alpha\beta\gamma}$ is the familiar antisymmetric tensor (Levi-Civita symbol).

To construct the second kind of operators we shall go back to Eqs. (14). Adding to the first of these $-i\epsilon K\mathbf{f}(\eta)$ and dividing by $-2i\epsilon K$ we have:

$$(i/2K\epsilon)(\epsilon^2\delta_{\alpha\beta}-k_\alpha k_\beta)u_\beta(\eta)+\tfrac{1}{2}f_\alpha(\eta) =\tfrac{1}{2}(\eta+1)f_\alpha(\eta). \quad (26_1)$$

The right-hand side has the factor 1 if $\eta=+1$ and zero if $\eta=-1$. Similarly we get from the second equation (14):

$$-(i/2K\epsilon)(K^2\delta_{\alpha\beta}+k_\alpha k_\beta)f_\beta(\eta)+\tfrac{1}{2}u_\alpha(\eta) =\tfrac{1}{2}(\eta+1)u_\alpha(\eta). \quad (26_2)$$

We introduce the abbreviations:

$$Y_{\alpha\beta}(\mathbf{k})=(i/2K\epsilon)(\epsilon^2\delta_{\alpha\beta}-k_\alpha k_\beta),$$
$$Z_{\alpha\beta}(\mathbf{k})=(1/2iK\epsilon)(K^2\delta_{\alpha\beta}+k_\alpha k_\beta). \quad (27)$$

III

(a) We shall now apply perturbation theory to (3) and (4). We develop the eigenvectors of the perturbed problem in terms of the unperturbed eigenvectors analogous to (22). Introducing these expressions into (3) and (4) we obtain:

$$\sum_{\eta,p}\int d\mathbf{k}\{c^{-1}A(\eta p\mathbf{k};t) +i(\hbar c)^{-1}A(\eta p\mathbf{k};t)\,V(\mathbf{x})\}\,{}_0\mathbf{F}(\eta p\mathbf{k};\mathbf{x}t)=0,$$

$$\sum_{\eta,p}\int d\mathbf{k}\{\qquad\qquad\qquad\qquad\}\,{}_0\mathbf{U}(\eta p\mathbf{k};\mathbf{x}t)=0.$$

The first of these integrals is now multiplied scalarly with $i^{-1}{}_0\mathbf{U}^*(\eta_1, p_1, \mathbf{k}_1;\mathbf{x},t)$ the second with $i_0\mathbf{F}^*(\eta_1, p_1, \mathbf{k}_1;\mathbf{x},t)$. Then the two are added and integrated over the $\mathbf{x}$ space. Applying to the term containing $\dot{A}$ the orthogonality conditions after the manner of Section 2(c) one finds

$$i\hbar\dot{A}(\eta_1 p_1\mathbf{k}_1;t)=\sum_{\eta,p}\int d\mathbf{k}A(\eta p\mathbf{k};t) \times(\eta_1 p_1\mathbf{k}_1|\,V|\,\eta p\mathbf{k})e^{ic(\eta_1\epsilon_1-\eta\epsilon)t}. \quad (28)$$

The matrix element of the perturbation potential V can easily be brought into connection with the usual matrix element of the Schrödinger equation:

$$(\eta_1 p_1\mathbf{k}_1|\,\tilde{V}\,|\,\eta p\mathbf{k})=(\eta_1 p_1\mathbf{k}_1|\,\tilde{\rho}\,|\,\eta p\mathbf{k})\cdot(\mathbf{k}_1|\,V|\mathbf{k}), \quad (29)$$

where

$$(\mathbf{k}_1|\,V|\mathbf{k})=(2\pi)^{-3}\int d\mathbf{x}\,V(\mathbf{x}) \times\exp[i(\mathbf{k}-\mathbf{k}_1\cdot\mathbf{x})]. \quad (30)$$

Before integrating (28) in the usual fashion, a definite initial condition has to be agreed upon. We shall assume that at $t=0$

(1) only particles of momentum $\hbar\mathbf{k}_0$, and
(2) only particles of positive energy, i.e., with $\eta_0=+1$ are present.
(3) For $p=+I$ the amplitude is to be $A_{+I}{}^{(0)}$, for $p=0$ $A_0{}^{(0)}$, and for $p=-I$ $A_{-I}{}^{(0)}$.

Thus the wave vectors of the incident particles are:

$$\mathbf{U}_{\text{prim}}=\sum_p A_p{}^{(0)}\,{}_0\mathbf{U}(1,p,\mathbf{k}_0;\mathbf{x},t),$$
$$\mathbf{F}_{\text{prim}}=\sum_p A_p{}^{(0)}\,{}_0\mathbf{F}(1,p,\mathbf{k}_0;\mathbf{x},t). \quad (31)$$

For $t=0$ the value of A is therefore

$$A(\eta,p,\mathbf{k};0)=\delta_{\eta,1}\delta(\mathbf{k}-\mathbf{k}_0)A_p{}^{(0)}. \quad (32)$$

To obtain the first approximation from (28) this condition is introduced into it on the right-hand term and the time integration performed. The result is:

$$\hbar cA^{(1)}(\eta,p,\mathbf{k};t)=-\eta\frac{e^{ic(\eta\epsilon_k-\epsilon_0)t}-1}{\eta\epsilon_k-\epsilon_0}(\mathbf{k}|\,V|\mathbf{k}_0) \cdot\sum_{p_0}A_{p_0}{}^{(0)}(\eta,p,\mathbf{k}|\,\tilde{\rho}\,|\,1,p_0,\mathbf{k}_0) \quad (33)$$

in which ϵ_0 is written instead of $\epsilon(\mathbf{k}_0)$. In the same standard manner the second approximation is obtained: The above first approximation is substituted into the right-hand side of (28). Integration gives:

$$(\hbar c)^2A^{(2)}(\eta,p,\mathbf{k};t)=\eta\sum_{\eta',p'}\eta'\int d\mathbf{k}' \times(\mathbf{k}|\,V|\mathbf{k}')(\mathbf{k}'|\,V|\mathbf{k}_0) \cdot\sum_{p_0}A_{p_0}{}^{(0)}(\eta,p,\mathbf{k}|\,\tilde{\rho}\,|\,\eta',p',\mathbf{k}') \times(\eta',p',\mathbf{k}'|\,\tilde{\rho}\,|\,1,p_0,\mathbf{k}_0)\cdot\Phi \quad (34_1)$$

with

$$\Phi(\eta,\eta';\epsilon,\epsilon',\epsilon_0;t)=\frac{e^{ic(\eta\epsilon-\epsilon_0)t}-1}{(\eta\epsilon-\epsilon_0)(\eta'\epsilon'-\epsilon_0)}-\frac{e^{ic(\eta\epsilon-\eta'\epsilon')t}-1}{(\eta\epsilon-\eta'\epsilon')(\eta'\epsilon'-\epsilon_0)}. \quad (34_2)$$

The thus-obtained values for $A(\eta, p, \mathbf{k}; t)$ should be put into (22) in order to furnish the eigenvectors of the perturbed problem in first or second approximation.

(b) It is, however, not necessary for our purpose to compute the eigenfunctions. All we are interested in is the number of particles $N(\mathbf{k}, t)$ scattered into the solid angle element $d\Omega$ between the time $t=0$ when the perturbation is switched on and the time $t=t$. This quantity is connected with the total number of scattered particles

$$\int d\Omega N(\mathbf{k},t)=\int d\mathbf{x} i^{-1}\{(\mathbf{U}^*\cdot\mathbf{F})-(\mathbf{F}^*\cdot\mathbf{U})\}$$

$$=\int d\mathbf{k}\sum_{\eta,\ p}\eta|A(\eta p\mathbf{k},t)|^2$$

according to (23). Since

$$d\mathbf{k}=d\Omega k^2 dk$$

one can drop the $d\Omega$ integration both sides. Finally the intensity measured at a distance R from the scatterer is given by:

$$J=\frac{1}{R^2}\frac{\partial N(\mathbf{k},t)}{\partial t}, \qquad =\frac{1}{R^2}\frac{\partial}{\partial t}\int_0^\infty k^2dk\sum_{\eta,\ p}\eta|A(\eta p\mathbf{k},t)|^2. \quad (35)$$

Substituting for A the sum of the expressions given by (33) and (34) one obtains J up to the second approximation.

IV

(a) In this section J will be evaluated in first approximation. Introducing (33) into (35) we have:

$$J^{(1)}=\frac{1}{R^2}\frac{\partial}{\partial t}\int_0^\infty k^2dk\sum_{\eta,\ p}\sigma(\eta p\mathbf{k};\mathbf{k}_0)\cdot 4\frac{\sin^2(c/2)(\eta\epsilon_k-\epsilon_0)t}{(\eta\epsilon_k-\epsilon_0)^2},$$

$$\sigma(\eta p\mathbf{k};\mathbf{k}_0)=(\hbar c)^{-2}\eta|(\mathbf{k}|V|\mathbf{k}_0)|^2\times\Big|\sum_{p_0}A_{p_0}^{(0)}(\eta p\mathbf{k}|\tilde{\rho}|1p_0\mathbf{k}_0)\Big|^2.$$

Only the term with $\eta=+1$ gives a resonance point within the range of the k integration. Since $\sigma(\cdots)$ is a slowly varying function of k it can be taken outside of the integral with $|\mathbf{k}_0|$ substituted for $|\mathbf{k}|$. The result is:

$$J^{(1)}=\frac{1}{R^2}\frac{2\pi}{\hbar^2c^2}\epsilon_0k_0|(\mathbf{k}|V|\mathbf{k}_0)|^2\sum_p\Big|\sum_{p_0}\times A_{p_0}^{(0)}(1p\mathbf{k}|\tilde{\rho}|1p_0\mathbf{k}_0)\Big|^2. \quad (36)$$

Although $|\mathbf{k}|^2=|\mathbf{k}_0|^2$, their directions differ of course. The term $|(\mathbf{k}|V|\mathbf{k}_0)|^2$ furnishes, for Coulomb interaction the well-known Rutherfordian angle function cosec^4 $(\vartheta/2)$.

(b) The other factor which we will call P contains the relativistic and polarization effects due to the Dirac-Proca equations. Written in a more detailed fashion it is according to (18):

$$P=i^{-2}\sum_p\sum_{p_0}\sum_{p_0'}A_{p_0}^{(0)*}A_{p_0'}^{(0)}\{u_\alpha^*(1p_0\mathbf{k}_0)f_\alpha(1p\mathbf{k})-f_\alpha^*(1p_0\mathbf{k}_0)u_\alpha(1p\mathbf{k})\}\cdot\{u_\beta^*(1p\mathbf{k})f_\beta(1p_0'\mathbf{k}_0)-f_\beta^*(1p\mathbf{k})u_\beta(1p_0'\mathbf{k}_0)\}. \quad (37)$$

The summation over p is easy to perform by means of the destruction operators of Section 2(e). For instance, the product of the first term of the first parenthesis with the first of the second parenthesis contains

$$\sum_p f_\alpha(1p\mathbf{k})u_\beta^*(1p\mathbf{k})=\sum_p\sum_\eta\eta\frac{\eta+1}{2}f_\alpha(\eta p\mathbf{k})u_\beta^*(\eta p\mathbf{k}).$$

Application of (26_1) from right to left and use of (27) make it possible to write this as:

$$\sum_{p,\eta}\eta[Y_{\alpha\gamma}(\mathbf{k})u_\gamma(\eta p\mathbf{k})+\tfrac{1}{2}f_\alpha(\eta p\mathbf{k})]u_\beta^*(\eta p\mathbf{k})=\tfrac{1}{2}i\delta_{\alpha\beta}.$$

The latter is in accord with the orthogonality conditions (21). Similarly from the second term of the first parenthesis and from the first term of the second parenthesis of (37) there arises:

$$\sum_p u_\alpha(1p\mathbf{k})u_\beta^*(1p\mathbf{k})=iZ_{\alpha\beta}(\mathbf{k}).$$

In this fashion (37) becomes:

$$P=\tfrac{1}{2}\sum_{p_0}|A_{p_0}{}^{(0)}|^2 + i^{-1}\sum_{p_0,p_0'}A_{p_0}{}^{(0)*}A_{p_0'}{}^{(0)}[\cdots], \quad (38)$$

$$[\cdots]=Y_{\alpha\beta}(\mathbf{k})u_\alpha{}^*(1p_0\mathbf{k}_0)u_\beta(1p_0'\mathbf{k}_0) - Z_{\alpha\beta}(\mathbf{k})f_\alpha{}^*(1p_0\mathbf{k}_0)f_\beta(1p_0'\mathbf{k}_0)$$

after account is taken of (18) and (20). Since the second term contains the arbitrary amplitude constants $A_{p_0}{}^{(0)}$, one has to spread out each term of the double sum into a sum of p and η in order to be able to apply the orthogonality conditions. Let us deal separately with terms where $p_0=p_0'$ and with those where $p_0\neq p_0'$.

(α) $p_0=p_0'$. The square bracket, the factor of $|A_{p_0}{}^{(0)}|^2$ has now to be treated with the operators D of Section 2(e). According to (24) and (26) we have:

$$u_\alpha{}^*(1p_0\mathbf{k}_0)u_\beta(1p_0\mathbf{k}_0)=D_{\beta\gamma}(p_0\mathbf{k}_0)\sum_{p,\eta}\eta u_\alpha{}^*(\eta p\mathbf{k}_0) \times\{Z_{\gamma\delta}(\mathbf{k}_0)f_\delta(\eta p\mathbf{k}_0)+\tfrac{1}{2}u_\gamma(\eta p\mathbf{k}_0)\}.$$

Treating the second term of $[\cdots]$ similarly, we finally obtain using (21):

$$[\cdots]=iD_{\alpha\beta}(p_0\mathbf{k}_0)\{Z_{\beta\gamma}(\mathbf{k}_0)Y_{\gamma\alpha}(\mathbf{k}) + Y_{\beta\gamma}(\mathbf{k}_0)Z_{\gamma\alpha}(\mathbf{k})\}. \quad (39)$$

From (25) and (27) we get as contribution to P in the case

$$p_0=p_0'=0: \quad |A_0{}^{(0)}|^2[1+(1/4K^2\epsilon^2)(k^4-(\mathbf{k}\cdot\mathbf{k}_0)^2)],$$
$$p_0=p_0'=\pm\mathrm{I}: \quad |A_{\pm 1}{}^{(0)}|^2[1+(1/8K^2\epsilon^2)(k^4-(\mathbf{k}\cdot\mathbf{k}_0)^2)]. \quad (40)$$

(β) $p_0=\mathrm{I}$, $p_0'=-\mathrm{I}$. By use of (27), (15), (16) and (19) the square bracket of (38) becomes:

$$i(k^2/4K^2\epsilon^2)(\mathbf{k}\cdot\mathbf{e}(-\mathrm{I},\mathbf{k}_0))^2.$$

As this expression no longer possesses cylindrical symmetry, it becomes convenient to work with the following rectangular components:

$$\mathbf{k}_0=(k,0,0),$$
$$\mathbf{e}(\mathrm{I},\mathbf{k}_0)=(0,2^{-\frac{1}{2}},i2^{-\frac{1}{2}}),$$
$$\mathbf{e}(-\mathrm{I},\mathbf{k}_0)=(0,2^{-\frac{1}{2}},-i2^{-\frac{1}{2}}),$$
$$\mathbf{k}=k(\cos\vartheta,\sin\vartheta\cos\varphi,\sin\vartheta\sin\varphi).$$

Then the contribution to P becomes:

$$A_1{}^{(0)*}A_{-1}{}^{(0)}\cdot\frac{k^4}{8K^2\epsilon^2}\sin^2\vartheta e^{-2i\varphi}. \quad (41)$$

(γ) Because of the Hermitian character of each term of (38) the contribution of $p_0=-\mathrm{I}$, $p_0'=+\mathrm{I}$ is

$$A_{-1}{}^{(0)*}A_1{}^{(0)}\cdot\frac{k^4}{8K^2\epsilon^2}\sin^2\vartheta e^{+2i\varphi}. \quad (42)$$

(δ) The contributions of terms one of whose p_0 numbers is zero vanish. Collecting terms from (40) to (42) the following expression for P is obtained:

$$P=\sum_{p_0}|A_{p_0}{}^{(0)}|^2+\frac{k^4}{4K^2\epsilon^2}\sin^2\vartheta \times\{|A_0{}^{(0)}|^2+\tfrac{1}{2}|A_1{}^{(0)}e^{i\varphi}+A_{-1}{}^{(0)}e^{-i\varphi}|^2\}. \quad (43)$$

(c) From this formula it is seen, that as long as the incident waves are solely longitudinal, or if transverse, strictly right or left circular, the inclusion of relativistic and spin terms gives only a ϑ effect which will be superimposed upon the Rutherfordian term $|(\mathbf{k}|V|\mathbf{k}_0)|^2$. But if both right and left circularly polarized waves are incident, i.e., if part of the incident wave is linearly polarized, an azimuth effect arises. This is seen when we put

$$A_1=A_L e^{-i\varphi_0}; \quad A_{-1}=A_L e^{i\varphi_0}.$$

Then P becomes, if we disregard longitudinal terms for the moment:

$$P=2A_L{}^2\{1+(K^4/4K^2\epsilon^2)\sin^2\vartheta\cos^2(\varphi-\varphi_0)\}.$$

The appearance of a polarization even in the first approximation seems surprising in com-

parison with the electron for which such effects according to Mott and Sauter do not appear until the second Born approximation. However, a certain similarity which the Dirac-Proca equations bear to the Maxwell equations and the vectorial character of the wave functions (as compared to the spinor character for $s=\frac{1}{2}$) may render our result more plausible.

The effect is of the fourth order in v/c for

$$k^4/K^2\epsilon^2=\beta^4/(1-\beta^2)$$

and is most pronounced at right angles to the direction of the incident particles.

The evaluation of the second approximation (34) and a closer discussion of the matrix element $(\mathbf{k}|\,V|\,\mathbf{k}_0)$ will be given in a later paper.

DECEMBER 1, 1938 PHYSICAL REVIEW VOLUME 54

The Vapor Pressure of Isotopes

KARL F. HERZFELD
The Catholic University, Washington, D. C.
AND
EDWARD TELLER
The George Washington University, Washington, D. C.

(Received October 10, 1938)

It is shown here that if all degrees of freedom in a vapor are nearly classical and if there is no association, the vapor pressure of the lighter isotope is always higher. If only the external degrees of freedom are nearly classical and the internal ones are in the lowest state, the coupling of purely harmonic vibrations has either no influence or tends probably to a further increase of the excess vapor pressure of the light isotope. Anharmonicity and the change in the van der Waals forces probably account for those cases in which the heavier isotope has the higher vapor pressure.

I. INTRODUCTION

THE difference in the physical and chemical equilibrium conditions for isotopes is a pure quantum phenomenon.[1] Classical theory would give no effects of the difference in mass. This latter only affects the statistical distribution of momenta.[2] In classical theory, however, the integration over the momenta is independent of the integration over the coordinates and influences the different states (e.g., vapor and liquid) in the same manner. Therefore it does not modify the equilibrium between two states. This suggests a systematic investigation of the isotope effect in different temperature regions.

II. LOW TEMPERATURES

At low temperatures the differences are most conspicuous. The adequate method which is in general use starts from the condition at absolute zero. The greatest effect in this region is due to the difference in zero point energy of the motion of the molecule as a whole in the field of its neighbors. Since the operator of the kinetic energy is positive definite the lowest quantum level of the lighter isotope will always be higher than that of the heavier isotope. As a consequence the lighter molecule will have a smaller heat of evaporation and therefore a greater vapor pressure.[3] The difference of entropy of translation and possibly rotation in the vapor state works in the opposite direction but has less importance in the low temperature region. All this has frequently been the subject of discussions.

III. HIGH TEMPERATURES

As mentioned before, there are no differences between isotopes at temperatures so high that

[1] We exclude cases in which gravity or centrifugal forces are of importance.

[2] The very small differences in potential energies for different isotopes shall be neglected except at the end of the paper.

[3] A possible exception might occur if one has association in the vapor phase connected with a very high intermolecular vibrational frequency.

classical treatment is justified. At temperatures not quite as high, the differences are due to small deviations from classical theory. It seems therefore to us that an expansion in powers of $\hbar$ (i.e., Planck's constant divided by 2π) in which the first term is given by classical theory is more appropriate than the method usually applied, which starts from considerations of the behavior near absolute zero. In fact if the latter method is applied the small differences in vapor pressure at high temperature result from cancelation of much bigger terms.

Expansions of the partition function in terms of $\hbar$ have been developed by Bloch[4] and by Wigner.[5] According to Wigner the first quantum correction is proportional to $\hbar^2$, the coefficient of $\hbar$ being zero.[6] In classical theory the partition function is given by

$$Z_{cl}=\int d\Omega \exp(-\epsilon/kT), \tag{1}$$

where $d\Omega$ signifies the volume element of the complete phase space and the energy ϵ is expressed in the form

$$\epsilon=\tfrac{1}{2}\sum_i p_i^2/m_i+V(x). \tag{2}$$

Here the p_i are the momenta, m_i the masses and V the potential energy, depending on all the coordinates x but not the momenta p or masses m. According to Wigner, $\exp(-\epsilon/kT)$ has to be replaced in first approximation of the quantum correction by

$$\exp(-\epsilon/kT)\left\{1+\frac{\hbar^2}{8(kT)^2}g_2\right\}, \tag{3}$$

where

$$g_2=\sum_i \frac{1}{m_i}\left[-\frac{\partial^2 V}{\partial x_i^2}+\frac{1}{3kT}\left(\frac{\partial V}{\partial x_i}\right)^2\right] +\sum_i\sum_j \frac{1}{3kT}\frac{p_i p_j}{m_i m_j}\frac{\partial^2 V}{\partial x_i \partial x_j}. \tag{4}$$

Therefore

$$Z=Z_{cl}+\frac{1}{8}\left(\frac{\hbar}{kT}\right)^2\int dx \exp\left(-\frac{V}{kT}\right) \times\int dp g_2 \exp\left(-\frac{1}{kT}\sum\frac{p^2}{2m}\right), \tag{5}$$

where dx, dp stand for the products of all dx_i, dp_i. It must be emphasized that in Wigner's method, the integrations over the p's and x's can be performed independently of each other, in spite of the fact that p and x cannot have definite values simultaneously in quantum theory.

We write

$$Z_{cl}=Z_{cl}'Z_{cl}''$$

where Z_{cl}' is the contribution from the kinetic, Z_{cl}'' that from the potential terms. Integrating over the momenta and performing a partial integration over the first terms of g_2, we get from (5)

$$Z=Z_{cl}'\left\{Z_{cl}''-\frac{1}{24}\left(\frac{\hbar}{kT}\right)^2 \times\left[\frac{1}{kT}\int dx\left(\exp-\frac{V}{kT}\right)\sum_i\frac{1}{m_i}\left(\frac{\partial V}{\partial x_i}\right)^2 +\text{surface terms}\right]\right\}. \tag{6}$$

The surface terms, resulting from the partial integration may be omitted, if we consider closed systems for which V becomes infinite at the boundary.

Therefore the partition function can be written

$$Z=Z_{cl}\left\{1-\frac{\hbar^2}{24(kT)^3}\left\langle\sum_i\frac{1}{m_i}\left(\frac{\partial V}{\partial x_i}\right)^2\right\rangle_{Av}\right\}, \tag{7}$$

where $\sum_i(1/m_i)(\partial V/\partial x)_i^2$ has to be averaged over the classical probability distribution in configuration space.

Part of the integration we have carried out has also been discussed by Kirkwood.[5] In comparing our expression with his formula (21) we can neglect his second term since it vanishes at high temperatures in an exponential manner.

[4] F. Bloch, Zeits. f. Physik **74**, 295 (1932).
[5] E. Wigner, Phys. Rev. **40**, 749 (1932). J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[6] For the special case of the harmonic oscillator, this is a well-known fact.

The factor T^{-3} in the correction term of (7) represents the actual temperature dependence only if

$$\left\langle \sum_i \frac{1}{m_i}\left(\frac{\partial V}{\partial x_i}\right)^2 \right\rangle_{Av}$$

is independent of the temperature. That is not the case if the atoms remain in the neighborhood of equilibrium positions, since under those conditions $(\partial V/\partial x)^2$ increases with T. With the help of a partial integration one can transform (7) into

$$Z = Z_{cl}\left\{1 - \frac{\hbar^2}{24(kT)^2}\left\langle \sum_i \frac{1}{m_i}\frac{\partial^2 V}{\partial x_i^2}\right\rangle_{Av}\right\}. \quad (7')$$

This formula applied to the harmonic oscillator yields the well-known result

$$Z = Z_{cl}\left\{1 - \frac{1}{24}\left(\frac{\hbar\omega}{kT}\right)^2\right\}. \quad (7'')$$

It can be seen from (7), that the quantum correction always diminishes the partition function and its absolute value increases with the mean square value of the forces and diminishing mass of the particles concerned. Thus, if we compare the liquid and vapor state, the quantum correction will be in general greater for the liquid on account of the stronger forces and the difference between gas and liquid is greater for the lighter isotope. Therefore the lighter isotope will have the higher vapor pressure. This has of course been known under the assumption of quasi-elastic forces, but has now been proved to be generally valid. It follows from our formulae that for rare gases the vapor pressure of the light isotope will be higher at sufficiently high temperatures. For molecules the equations deduced now can be applied only if all degrees of freedom behave classically. In such cases however the problem of comparing vapor pressures is frequently complicated by the fact that dissociation takes place. It is interesting to note that under the conditions which we are here discussing the dissociation into atoms will be always stronger for the lighter isotope. If one considers, therefore, the total vapor pressure of a molecule, dissociation into atoms can only cause a further relative increase of the vapor pressure of the lighter isotope. About the dissociation into radicals no general statements can be made.

IV. Intermediate Temperatures

In most liquids made up of molecules, the external degrees of freedom (i.e., the motion of a molecule as a whole in the field of its neighbors) behave in an almost classical way at temperatures where the internal degrees of freedom are still in the lowest state. Then considerations similar to those of the preceding section show that the influence of the external degrees of freedom will again be so as to increase the vapor pressure of the light isotope more than that of the heavy one above the classical value.

Nevertheless, a higher vapor pressure has been observed in many cases for the heavier isotope, e.g.,[7] in C_6D_6 compared to C_6H_6. This must be due to the influence of the internal vibrations, overcompensating the effect of the external degrees of freedom. Two ways have been proposed to explain this influence. The first explanation proceeds in the following manner.[8] It is well known that the frequencies of the internal vibrations are somewhat different in the gaseous and the liquid state. This difference will cause a difference in the heat of evaporation of the two isotopes. If, as is frequently observed in infra-red and Raman spectra, the frequency is lower in the liquid, the result would be a higher heat of evaporation and a lower vapor pressure of the lighter isotope. The second explanation[9] makes use of the different value of the van der Waals forces for the isotopes, which has its origin in the small difference of the polarizabilities and molecular volumes of the isotopes, which have been found experimentally.

With regard to the first explanation, one must keep in mind that the optically determined frequencies cannot be used directly to calculate average heats of evaporation. For the sake of simplicity we shall replace the liquid by a crystal.

[7] C. K. Ingold, C. G. Raisin and C. L. Wilson, J. Chem. Soc. 915 (1936).

[8] The effect in question has been pointed out by B. Topley and H. Eyring, J. Chem. Phys. **2**, 217 (1934). It has been used as an explanation by F. G. Brickwedde, Symposium on Recent Adv. in Physical Chemistry, A. A. A. S., Indianapolis, 1937.

[9] C. R. Bailey and B. Topley, J. Chem. Soc. 921 (1936).

For a fundamental frequency to be observed in the infra-red or Raman spectrum, neighboring molecules must vibrate practically in the same phase. This however is only one vibration among many others which have arbitrary values of the phase differences. These vibrations will have different frequencies; the optically observed frequency is probably either at the upper or the lower end of the frequency band. Thus the mean value of the frequencies in the condensed state which is the one to be used to determine effective zero-point energies, may differ considerably from the observed infra-red or Raman frequency.

In order to evaluate therefore the change of zero-point energy of the internal vibrations, it is necessary to consider theoretically the effect of the coupling of neighboring molecules. We shall discuss here only the sign of the change to be expected. It shall be assumed that all forces including those arising from neighboring molecules are harmonic. In view of the small changes expected we will be justified in applying perturbation theory and restricting ourselves to perturbations of the first order. In doing so, it will be sufficient to consider diagonal terms in the secular determinant, that is to consider the effect on the normal vibrations of one molecule, when it moves in the field of the neighboring molecules which are considered at rest.[10] Now the forces in the condensed state tend to hold the molecules in equilibrium configurations and therefore it is probable that the perturbing potential exerted on an atom by a neighboring molecule will have positive second derivatives. It is therefore to be expected that the mean value of the frequencies, and therefore the effective zero-point energy, will be higher in the condensed state and give rise to an additional excess of the vapor pressure of the light isotope. This last statement however is not to be considered as proved rigorously. It should be mentioned in particular, that according to Bauer and Magat[11] the anharmonicity of the internal vibrations and long range interactions (Coulomb forces) between atoms of different molecules may combine to cause an effective lowering of frequencies in the condensed state.

The second proposed explanation we have mentioned, i.e., the change of the van der Waals forces, seems to us to be the probable reason for the higher vapor pressure of the heavier isotope, whenever such an effect is observed for non-associated vapors. It can be seen from Bailey and Topley's paper[9] that the change in dispersion forces for benzene as calculated from the measured mole fractions and mole volumes would give rise to a greater difference in vapor pressure than has been observed. It is probable that the correct value would be obtained if the other phenomena discussed above could be taken into account and subtracted from the effect of the van der Waals forces.

The general explanation of the difference in van der Waals forces can be given by considering the difference of polarizabilities and dipole moments which are obtained by averaging over the zero-point vibrations of the internal degrees of freedom. As long as the vibration is strictly harmonic and the dipole moment and polarizability are strictly linear functions of all atomic displacements, the same mean values will be found for the two isotopes. However as soon as mechanical or electrical anharmonicities appear, a difference of the mean values and consequently of the van der Waals forces will be obtained, the sign of which depends on the specific properties of the molecules concerned.

[10] If we want to calculate the first-order perturbation for each mode of crystal vibration, those nondiagonal terms should also be considered, for which row and column belong to the same value of the unperturbed frequency. If however only mean values are wanted, one has to be concerned only with the *sum* of diagonal terms. Thus the Lorentz-Lorenz force will cause the splitting of an internal vibration into a band without affecting the mean value in the first order (see R. H. Lyddane and K. F. Herzfeld, Phys. Rev. **54**, 846 (1938)).

[11] E. Bauer and M. Magat, J. de Phys. **9**, 319 (1938).

On Thermal Dependence of Elasticity in Solids

LÉON BRILLOUIN
Collège de France, Physique Théorique, Paris, France
(Received October 4, 1938)

A theoretical study suggests a distinction between the elasticity coefficients for very high frequencies (of the elastic waves representing thermal agitation) and for low frequencies. These last coefficients should show a rapid decrease of rigidity when temperature increases, and the melting point can be said to represent the point at which the macroscopic rigidity becomes zero.

LAST year, while writing a book[1] on elasticity, in explaining the role of thermal agitation in solids, I tried to give a logical and systematic deduction of the theory of specific heat and thermal expansion of solids which obliged me to make some changes and improvements in the generally admitted theoretical scheme.

This brought me once more to the problem: How does temperature modify the elasticity coefficients of a solid body? This problem had troubled me for a long time and I had always felt some link was missing in the reasoning. In the many papers, books, and handbook articles on the subject, everyone, including myself, had started with the study of thermal agitation in an isotropic solid, thus analyzing thermal motion into elastic waves of the two classical longitudinal and transverse types. It was then assumed that the resulting formulas which had been developed for an isotropic solid were valid also for a solid distorted by stresses of any sort. A general thermodynamic theory of solids was built on this basis. This method is wrong because from a study strictly limited to isotropic solids one is not entitled to draw conclusions about what happens when the solid becomes anisotropic as a result of the action of external unsymmetrical forces.

In my book[1] (p. 344, 345) I merely noted this point and have only recently been able to put these general remarks into a more precise form. The necessary, though tiresome, calculations which I have performed entirely justify my point of view. Let us take an isotropic solid body, stretch it along the x direction by applying X forces; its elastic properties along the x direction now differ from those along the y and z direction. The solid has taken the symmetry of a uniaxial crystal. Calculation checks this prediction and shows that elastic waves are propagated in such a stretched solid body in a very peculiar way. There are no longer longitudinal and transverse waves but, for each direction of propagation, there are three different orthogonal polarizations, one of which is nearly longitudinal, another exactly transverse, and the third perpendicular to the first two and nearly transverse. To these three independent polarizations there correspond three different elastic wave velocities.

With this rigorous analysis of thermal agitation in a strained solid body it is possible to develop the whole thermodynamics of solids and to draw conclusions about the influence of temperature on elasticity coefficients. The general results are as follows.

(1) The elasticity coefficients which rule the propagation of hypersonic waves (as Raman calls them) are given by the derivatives of the elastic potential energy and will be affected only in an indirect way by the thermal expansion. They should accordingly show a slow decrease with increasing temperature.

(2) The macroscopic elasticity coefficients as measured in the propagation of acoustical or suprasonic waves are to be derived from the free energy instead of from the purely potential energy. These coefficients are directly influenced by thermal agitation. It is rather difficult to predict the variation of the λ-coefficient but the rigidity coefficient μ should show a very peculiar decrease with increasing temperature. This decrease, while slow at low temperatures, should become faster and faster as the temperature increases. For room temperature and for solids not too near their melting points the decrease is of the order of 50 to 100 times RT/V.

[1] L. Brillouin, *Les tenseurs en mécanique et en élasticité* (Masson, Paris, 1937), Chapter 12.

As the melting point is approached we may guess by extrapolation that the macroscopic rigidity will tend to zero while the microscopic rigidity remains finite. While melting, the body is unable to react to shear (on a macroscopic scale) but will still be able to propagate hypersonic transverse waves. Its specific heat, therefore, will remain about $3R$ just as for a solid body, and this is what really happens.

I wish to recall here an old paper of Sutherland[2] where the curious curve reproduced in

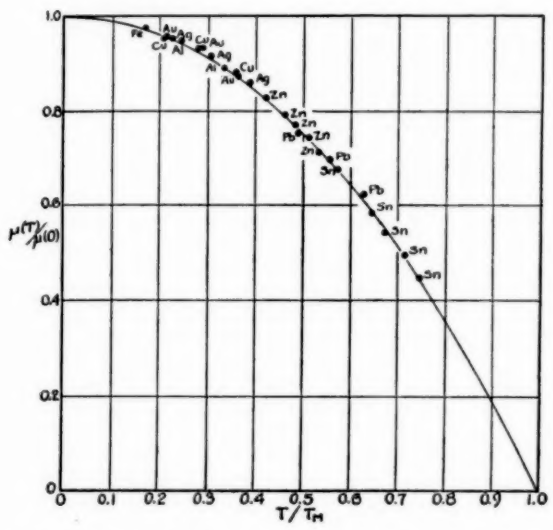

FIG. 1. Sutherland's curve (reference 2) showing the variation of $\mu(T)/\mu(O)$ with T/T_M. The curve is a portion of the parabola $y=1-x^2$.

Fig. 1 is to be found. In this curve the ordinates represent the ratio of $\mu(T)$ to $\mu(O)$ while the abscissas are the ratios of absolute temperature to melting point so that $x=1$ is the melting point. My father repeatedly emphasized that the melting point should be defined as the point at which the rigidity of a solid becomes zero. The preceding deductions seem to bring strong support to this prediction. It is, of course, difficult to imagine that the Sutherland curve should run smoothly up to the melting point since this would mean melting without latent heat of fusion. One must bear in mind, however, that the actual latent heats of fusion are always small compared to latent heats of vaporization. It is generally agreed that the liquid state is often much nearer to the solid state than to the gaseous state. Moreover, it must be remembered that isotropic solids are but a fiction which has been introduced only for simplification because the complete theory of elasticity would be too complicated in the case of crystals. We might guess that for a crystal the different rigidity coefficients would not tend all to zero at the same temperature. When one of these rigidity coefficients becomes zero the whole stability of the crystal is destroyed and this may involve a certain heat of transformation at the melting point. The complete study, a brief account of which is here given, will appear in the *Memorial des sciences mathematiques* (Gauthier-Villars, Paris) and is now in process of publication.

These theoretical remarks suggest experimental research on solids, and first of all, a systematic study of elasticity coefficients as functions of frequency and temperature. The temperature should be varied from very low temperatures up to the melting point. A connection should be sought with the elastic properties of the liquid above the melting point. Such measurements could be carried on by ultrasonic methods as already used by Balamuth, Rose, and Durand[3] and by Goens.[4] In a recent booklet[5] I have discussed the importance of these new methods of measurement which may lead to a better understanding of the properties of solids and the mechanism of fusion.

[2] Sutherland, Phil. Mag. **32**, 42 (1891).

[3] L. Balamuth, Phys. Rev. **45**, 715 (1934); **46**, 933 (1934). F. C. Rose, Phys. Rev. **49**, 50 (1936). M. A. Durand, Phys. Rev. **50**, 449 (1936).

[4] Goens, Physik. Zeits. **37**, 321 (1936).

[5] L. Brillouin, *La structure des corps solides* (A. S. I. No. 549, Hermann, Paris, 1937).

On a Minimum Property of the Free Energy

R. PEIERLS
The University, Birmingham, England
(Received September 22, 1938)

IT is well known that the eigenfunction ψ_1 of the lowest state of any system has the property of making the integral

$$\int \psi^* \mathbf{H} \psi d\tau \tag{1}$$

a minimum; the value of the integral is the corresponding eigenvalue E_1 of the Hamiltonian $\mathbf{H}$. It is also well known that this property leads to a powerful method of approximating, at least qualitatively, ψ_1 and E_1, by minimizing (1) among a restricted class of functions.

No similarly simple minimum property exists for the higher eigenvalues. One can obtain the nth eigenvalue by minimizing (1) with the subsidiary condition that ψ be orthogonal to $\psi_1, \psi_2, \cdots, \psi_{n-1}$. However, this procedure becomes progressively more cumbersome as n increases and besides it is of no use unless $\psi_1, \psi_2, \cdots, \psi_{n-1}$ are very exactly known.

For problems of thermal equilibrium one is often concerned with the free energy, rather than with the individual eigenvalues. It is the purpose of this note to draw attention to a simple minimum property of the free energy which may be considered as a generalization of the variation principle for the lowest eigenvalue.

The free energy has the following property: If $\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$ are an arbitrary set of orthogonal and normalized functions, and

$$H_{nn} = \int \varphi_n^* \mathbf{H} \varphi_n d\tau \tag{2}$$

is the expectation value of the Hamiltonian for the nth of them, then for any temperature T the function

$$\bar{F} = -kT \log \bar{P} = -kT \log \sum_n e^{-H_{nn}/kT}, \tag{3}$$

which *would* represent the free energy if the H_{nn} *were* the true eigenvalues, is higher than the true free energy[1]

$$F_0 = -kT \log P_0 = -kT \log \sum_n e^{-E_n/kT}. \tag{4}$$

[1] For $T=0$ this theorem evidently reduces to the variation principle for the lowest eigenvalue E_1.

This is equivalent to saying that the partition function, as formed by means of the expectation values H_{nn}:

$$\bar{P} = \sum_n e^{-H_{nn}/kT} \tag{5}$$

is less than the true partition function

$$P_0 = \sum_n e^{-E_n/kT}.$$

In this form our theorem is a special case of the more general statement that if $f(E)$ is a function with the properties

$$df/dE < 0, \quad d^2f/dE^2 > 0, \tag{6}$$

the expression

$$f = \sum_n f(H_{nn})$$

is less than

$$f_0 = \sum_n f(E_n).$$

(We shall always assume that the latter is finite.)

We prove this first for the partial sum consisting of the first N terms. This partial sum,

$$\bar{f}_N = \sum_{n=1}^{N} f(H_{nn}) \tag{7}$$

is bounded, as it cannot exceed the value $N \cdot f(E_1)$, according to (6). There must therefore be a greatest value of $\bar{f}_N$ and we shall show that this cannot be assumed for any set of functions other than $\psi_1, \psi_2, \cdots, \psi_N$.

Suppose that the set $\varphi_1, \varphi_2, \cdots, \varphi_N$ yields the maximum value of $\bar{f}_N$ and suppose that these functions are not all eigenfunctions. Then they do not make the energy diagonal, and hence at least one of the nondiagonal elements, say H_{mn}, with $m \leqq N$, must differ from zero. We must then distinguish three cases,

(a) $n > N$,

(b) $n \leqq N$, $H_{mm} \neq H_{nn}$,

(c) $n \leqq N$, $H_{mm} = H_{nn}$.

Case a. In this case we do not spoil the orthogonality of our set if we replace φ_m by

$$\varphi_m' = \varphi_m + \epsilon\varphi_n \tag{8}$$

with infinitely small ϵ. This leads to replacing the expectation value H_{mm}, up to first order, by

$$H_{mm}' = H_{mm} + (\epsilon H_{mn} + \epsilon^* H_{nm})$$

and the corresponding change in $\bar{f}_N$ is:

$$f_N' - \bar{f}_N = (df/dE)_{H_{mm}}(\epsilon H_{mn} + \epsilon^* H_{nm}).$$

It is obvious that this can be made positive by suitably choosing the sign of ϵ, and this is in contradiction with the assumption that our original set gave the maximum possible value of $\bar{f}_m$.

Case b. Here both φ_m and φ_n occur among the first N functions, and we replace them by

$$\varphi_m' = \varphi_m + \epsilon\varphi_n, \quad \varphi_n' = \varphi_n - \epsilon^*\varphi_n. \tag{9}$$

This changes H_{mm} and H_{nn} into

$$\begin{aligned} H_{mm}' &= H_{mm} + (\epsilon H_{mn} + \epsilon^* H_{nm}); \\ H_{nn}' &= H_{nn} - (\epsilon H_{mn} + \epsilon^* H_{nm}), \end{aligned} \tag{10}$$

except for second-order terms, and hence the change in $\bar{f}_N$ is, to first order:

$$f_N' - \bar{f}_N = \{(df/dE)_{H_{mm}} - (df/dE)_{H_{nn}}\} \times (\epsilon H_{mn} + \epsilon^* H_{nm}).$$

Since because of (6) the condition $H_{mm} \neq H_{nn}$ implies

$$(df/dE)_{H_{mm}} - (df/dE)_{H_{nn}} \neq 0,$$

(11) can be made positive by suitably choosing ϵ, and we arrive at the same contradiction.

Case c. If $H_{mm} = H_{nn}$, our substitution (9) produces no first-order change in $\bar{f}_N$. Proceeding to terms of the second order (for which the change of normalization has to be taken into account) one easily verifies that because of the equality of the diagonal elements, (10) gives the change of the expectation values correctly to second order, and hence the change in f_N becomes

$$f_N' - \bar{f}_N = (d^2f/dE^2)_{H_{mm}}(\epsilon H_{mn} + \epsilon^* H_{nm})^2.$$

This, as the square of a real quantity, is always positive.

We have thus, in all three cases, arrived at a contradiction to the assumption that a set of functions which are not all eigenfunctions gave the maximum value of $\bar{f}_N$. Hence the maximum must belong to a set consisting of N eigenfunctions, and it will then obviously be reached for the first N of them, and hence be equal to f_{N0}.

Consider now the infinite sum $\bar{f}$. If, for a suitable set of functions, it could reach a value exceeding f_0, there would have to exist a number N such that

$$f_N > f_{N0},$$

since for $N \to \infty$, $\bar{f}_N \to \bar{f}$ and $f_{N0} \to f_0$. This, however, we have just shown to be impossible.

So far we have assumed our set of functions to be complete. Our theorem holds *a fortiori* for an incomplete orthogonal set, since an incomplete set can always be thought of as formed by omitting some of the functions from a complete set; and by omitting terms the value of the sum (7) is reduced.

If we insert for the general function f:

$$f(E) = e^{-E/kT},$$

we obtain the above theorem about the free energy, but by inserting other functions it is easy to see that the theorem also applies to the free energy of an assembly of independent particles satisfying Bose-Einstein or Fermi-Dirac statistics.

It is hoped that an illustration of the application of this theorem will soon be published by Mr. F. Hoyle.

The Electrodynamics of Material Media

CARL ECKART
Ryerson Laboratory, The University of Chicago, Chicago, Illinois
(Received September 9, 1938)

The general equations of a gauge invariant, classical theory of the electrodynamics of material media are obtained. The gauge invariance is insured by taking the equations

$$\nabla\cdot\mathbf{D}=\rho, \quad \nabla\times\mathbf{H}-\mathbf{D}'=\mathbf{J}$$

as conditions auxiliary to the variation principle

$$\delta\int\int\int\int\{L+\Sigma_n\theta_n[N_n'+\nabla\cdot(N_n\mathbf{V}_n)]dvdt\}=0.$$

The Lagrangian function, L, depends on $\mathbf{D}$, $\mathbf{H}$, N_n, $\mathbf{V}_n$, θ_n and possibly their derivatives; here N_n is the numerical density of atoms in the state n, $\mathbf{V}_n$ their macroscopic or average velocity, and θ_n is a variable that functions as the velocity potential in some cases and has the dimensions of action. The electromagnetic potentials enter the theory as Lagrangian multipliers only.

It is shown that if there is only one state and

$$L=\tfrac{1}{2}Nm\mathbf{V}^2-(\hbar^2/8m)(\nabla N)^2/N+\tfrac{1}{2}(\mathbf{H}^2-\mathbf{D}^2),$$

then the Schrödinger wave equation is obtained on making the substitution

$$\psi=N^{\frac{1}{2}}\exp(-i\theta/\hbar).$$

If the atoms are stationary (so that terms in $\mathbf{V}_n$ may be neglected), and

$$L=\Sigma_nN_nW_n+\mathbf{D}\cdot\mathbf{P}+\tfrac{1}{2}(\mathbf{H}^2-\mathbf{D}^2),$$

where W_n is the energy of the nth state, and

$$\mathbf{P}=\Sigma\Sigma_{(m,\,n)}(N_mN_n)^{\frac{1}{2}}\mathbf{P}_{mn}\cos[(\theta_m-\theta_n)/\hbar+\alpha_{nm}]$$

is the polarization of the medium, an adequate theory of dispersion results. However, the spontaneous transitions are not correctly accounted for by the equations.

If the electromagnetic fields are neglected and

$$L=\tfrac{1}{2}Nm\mathbf{V}^2-U(N),$$

the equations are those for the irrotational motion of a gas, with θ/m as the velocity potential.

INTRODUCTION

ELECTRODYNAMICS has never been developed in a manner analogous to analytic mechanics. Some years ago, G. Mie[1] made an important attempt in this direction, and deduced the equations of his theory from a single Lagrangian function. One consequence of this attempt was a formulation of the requirement of gauge invariance[2] by H. Weyl; unfortunately, Mie's theory did not meet this requirement, nor can Weyl's geometrical resolution of the difficulty be considered as satisfactory at the present time. It is the purpose of this paper to indicate the possibility of a classical theory that is in some respects similar to Mie's and satisfies the requirement of gauge invariance.

In addition to its use of a Lagrangian function, analytic mechanics is characterized by the use of a great variety of coordinates. In any special case, these are chosen so as to make the introduction of empirical data into the calculations as simple as possible. No formal attempt to define the coordinates in terms of atomic quantities is made, nor to deduce their equations of motion from those of the ultimate particles. On the other hand, the terms and suggestions of atomic theory are often used freely in the informal explanations that accompany the mathematical theory.

A similar procedure was suggested by W. Heisenberg[3] as a basis for the development of quantum theory. It is interesting to note the close relation between the initial stages of the development of quantum theory and the classical theory outlined below. However, this does not make the later parts of quantum theory superfluous; instead, the need for them becomes more apparent.

THE GENERAL EQUATIONS

The medium may be composed of various kinds of particles, capable of existing in different states. For simplicity, only a finite number of states will be considered. Let N_n be the numerical density or concentration of particles in the state[4] n, and $\mathbf{V}_n$ their macroscopic or average velocity. Another scalar function of position and time, θ_n,

[1] G. Mie, Ann. d. Physik **37**, 511 (1912); **39**, 1 (1912); **40**, 1 (1913).
[2] H. Weyl, Sitzungsber. d. Preuss. Akad. (1918).
[3] W. Heisenberg, Zeits. f. Physik **33**, 879 (1925).
[4] For the present, there is no need to distinguish between the various kinds of particles. The notation can thus be simplified by avoiding all reference to the kinds of particles and numbering the states in some consecutive order.

will also be associated to the state n. This variable is analogous to the velocity potential of the irrotational motion of a fluid; it is also analogous to Dirac's[6] angle variable that is canonically conjugate to the square root of the total number of particles.

These variables will not, in general, be sufficient to describe the state of the medium: Such other quantities as the electromagnetic field, the gravitational field, and the temperature, are also needed. For the present, all of these except the electromagnetic field will be ignored, and this will be described by the vectors $\mathbf{D}$ and $\mathbf{H}$, that satisfy the Maxwell equations

$$\nabla\cdot\mathbf{D}=\rho, \quad \nabla\times\mathbf{H}-\mathbf{D}'=\mathbf{J}. \tag{1}$$

Here ρ and $\mathbf{J}$ are the electric charge and current densities, so that if e_n is the net charge carried by a particle of kind and state n, then

$$\rho=\Sigma_n N_n e_n, \quad \mathbf{J}=\Sigma_n N_n e_n \mathbf{V}_n. \tag{2}$$

The dynamical properties of the medium are supposed to be summarized by a variation principle involving the Lagrangian, L, which depends on N_n, $\mathbf{V}_n$, θ_n, $\mathbf{D}$ and $\mathbf{H}$. It may also depend on the space and time derivatives of these quantities, but for simplicity it will be supposed independent of all derivatives except ∇N_n. The variation principle is

$$\delta\int\int\int\int\{L+\Sigma_n\theta_n \times[N_n'-\nabla\cdot(N_n\mathbf{V}_n)]\}dvdt=0, \tag{3}$$

subject to Eqs. (1) and (2) as auxiliary conditions. If L has the dimension of energy per unit volume, this equation shows that θ_n has the dimension of action. As is customary in physics, not much will be said concerning the boundary of the region of integration, nor about the conditions there; however, there is good reason to believe that this constitutes an unnecessary limitation upon the power of the theory.

The Eulerian equations of this problem will be obtained by the method of Lagrangian multipliers. Let ϕ and $-\mathbf{A}$ be the multipliers associated to the Eq. (1): Then the variational integral becomes

$$\delta\int\int\int\int\{L+\Sigma_n\theta_n[N_n'+\nabla\cdot(N_n\mathbf{V}_n)] +\phi[\nabla\cdot\mathbf{D}-\rho]-\mathbf{A}\cdot[\nabla\times\mathbf{H}-\mathbf{D}'-\mathbf{J}]\}dvdt=0;$$

ρ and $\mathbf{J}$ are treated as abbreviations for the sums in Eq. (2). It is convenient to use the following notation: if $\mathbf{Q}=\mathbf{i}Q_x+\mathbf{j}Q_y+\mathbf{k}Q_z$ is any vector, then

$$(\partial L/\partial\mathbf{Q})=\mathbf{i}(\partial L/\partial Q_x)+\mathbf{j}(\partial L/\partial Q_y)+\mathbf{k}(\partial L/\partial Q_z).$$

The equations obtained from the variation of $\mathbf{D}$ and $\mathbf{H}$ are

$$(\partial L/\partial\mathbf{D})-\nabla\phi-\mathbf{A}'=0,$$
$$(\partial L/\partial\mathbf{H})-\nabla\times\mathbf{A}=0.$$

With the abbreviations

$$\mathbf{E}=-(\partial L/\partial\mathbf{D}), \quad \mathbf{D}=(\partial L/\partial\mathbf{H}), \tag{4}$$

they become

$$\mathbf{E}=-\nabla\phi-\mathbf{A}', \quad \mathbf{D}=\nabla\times\mathbf{A} \tag{5}$$

so that ϕ and $\mathbf{A}$ are seen to be the electromagnetic potentials. They do not necessarily satisfy Maxwell's auxiliary equation. They may be eliminated from Eq. (5) by differentiation, and there result the remaining Maxwell equations

$$\nabla\cdot\mathbf{B}=0, \quad \nabla\times\mathbf{E}+\mathbf{B}'=0. \tag{6}$$

The equation obtained by varying θ_n is

$$(\partial L/\partial\theta_n)+N_n'+\nabla\cdot(N_n\mathbf{V}_n)=0, \tag{7}$$

so that

$$G=-(\partial L/\partial\theta_n) \tag{8}$$

is the net rate (per unit of time and volume) at which particles are entering the state n. The result of the variation of N_n and $\mathbf{V}_n$ is

$$(\partial L/\partial N_n)-\nabla\cdot(\partial L/\partial\nabla N_n)-\theta_n' -\mathbf{V}_n\cdot\nabla\theta_n-e_n\phi+e_n\mathbf{V}_n\cdot\mathbf{A}=0, \tag{9}$$

$$(\partial L/\partial\mathbf{V}_n)-N_n\nabla\theta_n+N_n e_n\mathbf{A}_n=0. \tag{10}$$

The Eqs. (9) and (10) can be combined to give

$$\theta_n'=[(\partial L/\partial N_n)-\nabla\cdot(\partial L/\partial\nabla N_n)] -[\mathbf{V}_n\cdot(\partial L/\partial\mathbf{V}_n)]/N_n-e_n\phi, \tag{11}$$

$$\mathbf{V}_n\cdot\nabla\theta_n=[\mathbf{V}_n\cdot(\partial L/\partial\mathbf{V}_n)]/N_n-e_n\mathbf{V}_n\cdot\mathbf{A}, \tag{12}$$

which are frequently convenient.

Any pair of the Eqs. (10) and (11) may be used to eliminate ϕ and $\mathbf{A}$ from any other equa-

[6] P. A. M. Dirac, *Quantum Mechanics* (Oxford Press, 1935), Section 62 (second edition).

tions in which they may occur. This remark is the proof of the gauge invariance of the present theory.

After a somewhat tedious calculation, the energy equation is obtained as

$$\{\mathbf{B}\cdot\mathbf{H}-L+\sum_n \mathbf{V}_n\cdot(\partial L/\partial \mathbf{V}_n)\}'+\nabla\cdot\{\mathbf{E}\times\mathbf{H} +\sum_n[N_n'(\partial L/\partial\nabla N_n)+\mathbf{V}_n\mathbf{V}_n\cdot(\partial L/\partial\mathbf{V}_n) -N_n\mathbf{V}_n\{(\partial L/\partial N_n)-\nabla\cdot(\partial L/\partial\nabla N_n)\}]\}, \quad (13)$$

and the momentum equation (x component only) as

$$\{(\mathbf{D}\times\mathbf{B})_x+\sum_n N_n(\partial\theta_n/\partial x)\}'+(\partial/\partial x)\{\mathbf{E}\cdot\mathbf{D} +L-\sum_n N_n[(\partial L/\partial N_n)-\nabla\cdot(\partial L/\partial\nabla N_n)]\} +\nabla\cdot\{-\mathbf{D}E_x-\mathbf{B}H_x-\sum_n[N_n\mathbf{V}_n(\partial\theta_n/\partial x) -(\partial N_n/\partial x)(\partial L/\partial\nabla N_n)]\}=0. \quad (14)$$

The calculation for Eq. (13) is as follows: a in the expression for L', the terms $\mathbf{B}\cdot\mathbf{H}'-\mathbf{E}\cdot\mathbf{D}'$ are transformed as in the usual proof of Poynting's theorem; b the term in $\nabla N_n'$ is transformed into a divergence and a term proportional to N_n'; c the partial derivatives of L are eliminated by Eqs. (7), (9) and (10); d on collecting terms, the Eqs. (11) and (12) yield Eq. (13). The calculation for Eq. (14) is similar.

Special Examples

1.

As a particular case of these equations, let there be only one kind of particle, capable of existing in only one state. Then, if

$$L=\tfrac{1}{2}Nm\mathbf{V}^2-(\hbar^2/8mN)(\nabla N)^2+\tfrac{1}{2}(\mathbf{H}^2-\mathbf{D}^2), \quad (15)$$

the Eqs. (7), (10) and (11) become[6]

$$N'+\nabla\cdot(N\mathbf{V})=0,$$
$$m\mathbf{V}=\nabla\theta-e\mathbf{A},$$
$$\theta'=-\tfrac{1}{2}m\mathbf{V}^2+(\hbar^2/4m)[\tfrac{1}{2}(\nabla N/N)^2 +\nabla\cdot(\nabla N/N)]-e\phi. \quad (16)$$

The substitution

$$\psi=N^{\frac{1}{2}}\exp(-i\theta/\hbar) \quad (17)$$

reduces these to Schrödinger's equation

$$-ih\psi'=(1/2m)[-ih\nabla+e\mathbf{A}]^2\psi+e\phi\psi. \quad (18)$$

[6] Since L is independent of θ, the latter may here be considered as a Lagrangian multiplier associated to the auxiliary condition $N'+\nabla\cdot(N\mathbf{V})=0$.

By introducing two and four states, respectively, and a suitable L, the Pauli and Dirac equations can presumably be obtained. In these cases, the Lagrangian must depend explicitly on the variables θ_n.

In considering the significance of this example, there is one point that must not be overlooked. The potentials ϕ and $\mathbf{A}$ that enter into Eq. (18) are not independent of ψ. Eq. (18) is only one consequence of the Eq. (16), and Eqs. (1), (2), (4) and (5) are to be solved simultaneously with Eq. (16). In particular, it can be shown that the assumption

$$\phi=-e/r, \quad \mathbf{A}=0$$

is inconsistent with these other equations. Now, it is only because of this special assumption regarding ϕ and $\mathbf{A}$ that Schrödinger obtained agreement with the empirical data on the spectrum of hydrogen[7]: Consequently it is not correct to say that the present theory contains the Schrödinger theory as a special case. This conclusion applies with redoubled force when polyelectronic atoms are considered.

2.

Equally important is the possibility of obtaining an analytic dispersion theory. For this purpose, the atoms may be supposed fixed, and the terms in $\mathbf{V}_n$ ignored. The states are to be interpreted as the discrete energy states (energy $=W_n$) of a single kind of atoms. The atoms may be supposed neutral, so that $\rho=0$, $\mathbf{J}=0$, and the polarization of the medium is assumed to be

$$\mathbf{P}=\sum\sum_{(m,\,n)}(N_mN_n)^{\frac{1}{2}}\mathbf{P}_{mn} \times\cos[(\theta_m-\theta_n)/\hbar+\alpha_{mn}], \quad (19)$$

where $\mathbf{P}_{nm}\exp(i\alpha_{nm})$ is the matrix element of the dipole moment. If

$$L=\Sigma_n N_nW_n+\mathbf{D}\cdot\mathbf{P}+\tfrac{1}{2}(\mathbf{H}^2-\mathbf{D}^2), \quad (20)$$

Eqs. (4) become

$$\mathbf{E}=\mathbf{D}-\mathbf{P}, \quad \mathbf{B}=\mathbf{H}, \quad (21)$$

and Eqs. (7) and (9) become, after neglecting terms in $\mathbf{V}_n$,

$$N_n'+\mathbf{D}\cdot(\partial\mathbf{P}/\partial\theta_n)=0,$$
$$-\theta_n'+W_n+\mathbf{D}\cdot(\partial\mathbf{P}/\partial N_n)=0. \quad (22)$$

[7] E. Schrödinger, Ann. d. Physik (4) **79**, 361, 489 (1926).

A first approximation to the solution of these equations is sufficient to give an adequate account of dispersion phenomena; a second approximation will probably give a correct account of the phenomena of absorption and forced emission,[8] but it will be necessary to add new terms to the Lagrangian if the phenomenon of spontaneous emission is to be correctly included in this approximation. Presumably, with each successive approximation, new terms will be needed in the Lagrangian in order to bring the theory into agreement with the empirical data. If this series of terms converged to some simple function, the situation would be satisfactory; it is not possible to foresee that this will be the case.

It would be interesting to investigate the possibility of replacing Eq. (19) by another; the factors $(N_n N_m)^{\frac{1}{2}}$ are well known from the quantum theory, but seem somewhat strange from the classical point of view, unless the exponent arises from a root-mean-square average based on some statistical theory. It may be that some other dependence on the concentrations will be equally (or even, more) satisfactory.The success of such an investigation seems somewhat doubtful to the author (for reasons similar to those discussed at the end of Example 1), but it should be undertaken in any case; it lies beyond the scope of this outline.

3.

The equations for the irrotational motion of a gas are also a special case of the foregoing. The terms in the electromagnetic fields are unimportant in this connection and may be ignored; the Lagrangian is

$$L=\tfrac{1}{2}Nm\mathbf{V}^2-U(N). \tag{23}$$

Since it is independent of the variable θ, the total number of atoms remains constant. The Eqs. (9) and (10) become

$$\begin{aligned}\tfrac{1}{2}m\mathbf{V}^2-(\partial U/\partial N)-\theta'-\mathbf{V}\cdot\nabla\theta=0,\\ m\mathbf{V}-\nabla\theta=0,\end{aligned} \tag{24}$$

so that θ/m is the ordinary velocity potential in this case. The Eqs. (24) are not Euler's equations of hydrodynamics; the latter are essentially Eq. (14), which reduces to

$$\begin{aligned}(NmV_x)'+(\partial/\partial x)[N(\partial U/\partial N)-U]\\ +\nabla\cdot[Nm\mathbf{V}V_x]=0,\end{aligned} \tag{25}$$

in this case. The pressure is thus related to the function U by the equation

$$p=N(\partial U/\partial N)-U. \tag{26}$$

If $U=aN^{\gamma+1}$, γ being the ratio of the gas constant to the specific heat at constant volume, then $p=a\gamma N^{\gamma+1}$ and the motion is adiabatic. If $U=NRT\log N$, where R is the gas constant per molecule and $T=$constant, then $p=NRT$, and the motion is isothermal, with temperature T.

Conclusion

The "analytic" or phenomenologic electrodynamics outlined here cannot pretend to compete with quantum electrodynamics in the domain proper to the latter. However, the former should not be regarded only as a branch of pure mathematics that happens to use the terms of physics. There are undoubtedly many macroscopic phenomena whose laws are special cases of the equations developed above.

Neither is this theory merely the Maxwell-Lorentz theory in a new form. The latter contains Eqs. (1), (2), and (6), together with analogs to Eqs. (4), (13), and (14); it does not contain analogs to Eqs. (9) and (10).

It is my privilege to dedicate this paper to Professor Arnold Sommerfeld, as a greeting on the occasion of his seventieth birthday.

[8] Except for notation, these equations are very similar to those studied by Dirac, Proc. Roy. Soc. **114**, 710 (1927). The statements in the text are based on this similarity.

On Asymptotic Series for Functions in the Theory of Diffraction of Light

W. PAULI
The Institute of Physics of the Eidgenössische Technische Hochschule, Zürich, Switzerland
(Received October 3, 1938)

For Sommerfeld's exact solution of the problem of the diffraction of light by a wedge, limited by two perfectly reflecting planes, new asymptotic formulas are given; these also hold near one of the boundaries between light and shadow, at a distance from the diffracting body which is large compared with the wave-length of light.

1. INTRODUCTION. STATEMENT OF THE PROBLEM

AS is well known, A. Sommerfeld, who is now celebrating his seventieth birthday, was the first to succeed in establishing an exact solution of the wave equation for a particular class of two-dimensional problems of diffraction.[1] The problem in question is the diffraction by a wedge, limited by two perfectly reflecting planes. The diffraction by a half-plane is contained therein as a particular case, when the angle θ between the limiting planes of the wedge tends to zero.

For the sake of simplicity we shall assume that the monochromatic source of light is at infinity and that the direction of propagation of the incident plane wave is perpendicular to the edge of the wedge, which edge may coincide with the z axis of the coordinate system. By splitting off the factor $\exp(+i\omega t)$ from all field quantities and taking into consideration, that all field quantities become independent of z, it is easy to reduce the problem to the solution of the two-dimensional scalar wave equation:

$$(\partial^2/\partial x^2+\partial^2/\partial y^2)u+k^2u=0 \qquad (1)$$

with $k=\omega/c$. According to the two possible states of the polarization of the incident wave (whether the electric vector oscillates perpendicularly to the plane of incidence (A), or within this plane (B)) one has to fulfill the boundary conditions:

$$\text{(A)}\quad u=0; \qquad \text{(B)}\quad \partial u/\partial n=0,$$

where n means as usual the direction of the normal upon the surface of the wedge. In either case the boundary condition of a perfectly reflecting mirror calls for the vanishing of the tangential components of the electric field strength.

In the case (A) the nonvanishing components of E and H are H_x, H_y, E_z, and as a consequence of Maxwell's equations one can put

$$E_z=u, \quad H_x=(i/k)(\partial u/\partial y), \quad H_y=-(i/k)(\partial u/\partial x).$$

The boundary condition is fulfilled if $u=0$. In the case (B) the components E_x, E_y, H_z are different from zero and one can put

$$E_x=-(i/k)(\partial u/\partial y), \quad E_y=(i/k)(\partial u/\partial x), \quad H_z=u.$$

The postulate, that the tangential components of E vanish, is now equivalent to $\partial u/\partial n=0$.

Now we introduce cylindrical coordinates defined by

$$x=r\cos\psi, \quad y=r\sin\psi$$

in such a way, that the origin of the coordinate system coincides with the edge of the wedge and that the values $\psi=0$ and $\psi=2\pi-\theta$ of the azimuth ψ correspond to the boundary planes of the wedge. If $\psi=\psi_0$ is the azimuth of the radius vector drawn from the origin to the light source at infinity (so that $0\leqq\psi_0\leqq 2\pi-\theta$), the incident wave is represented by

$$u_0=e^{ikx\cos\psi_0}, \qquad (2)$$

where as always in the sequel the factor $e^{+i\omega t}$ is omitted and the amplitude of u_0 is normalized to unity. (Compare Fig. 1. The arrow indicates the direction of the incident wave and the shaded domain the shadow according to geometrical optics.)

[1] A. Sommerfeld, Math. Ann. **45**, 263 (1894) and **47**, 317 (1896). For the case of the wedge, H. M. Macdonald, *Electric Waves* (Cambridge, 1902), p. 186, and H. S. Carslaw, Proc. Lond. Math. Soc. **18**, 291 (1919) obtained an essential simplification of the solution. We follow here also, in the notation, a summarizing article written by Sommerfeld himself in the book edited by P. Frank and R. V. Mises, *Differential- u. Integralgleichungen der Physik*, 2nd edition (1935), Vol. 2, Chap. 20.

By putting

$$u=v(r,\psi-\psi_0)\pm v(r,\psi+\psi_0), \tag{3}$$

which corresponds to the separation of the total field into the incident and the reflected wave, the solution of the problem is reduced to the function $v(r,\varphi)$ of Sommerfeld. This is a solution of the wave equation (1), which has several branches in the (x,y) plane; it has the period $2\pi n$ in the azimuth φ, but is a one valued function of the variable $\cos\varphi/n$. One has therefore

$$v(r,\varphi+2\pi n)=v(r,\varphi), \tag{4a}$$

$$v(-\varphi)=v(\varphi),\quad v(\pi n-\varphi)=v(\pi n+\varphi). \tag{4b}$$

If the half-period πn of the function $v(r,\varphi)$ coincides with the azimuth $2\pi-\theta$ of the second limiting surface of the wedge

$$\pi n=2\pi-\theta, \tag{5}$$

the boundary conditions (A) or (B) are fulfilled on both surface planes of the wedge, according to whether one assumes the upper or the lower sign in Eq. (3).

We restrict ourselves to the case

$$n>1,\quad \theta<\pi \tag{6}$$

where, however, it is in no way assumed that n is an integer. Moreover we introduce the abbreviation

$$\rho=kr \tag{7}$$

for the "numerical distance." The following further conditions have to be fulfilled, in order

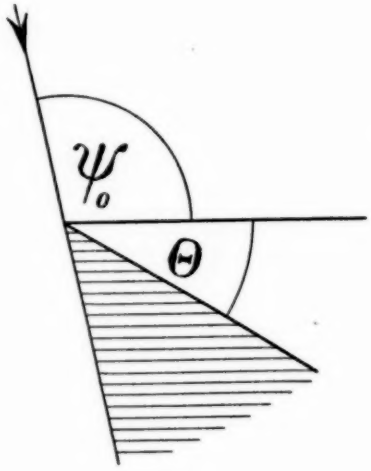

FIG. 1.

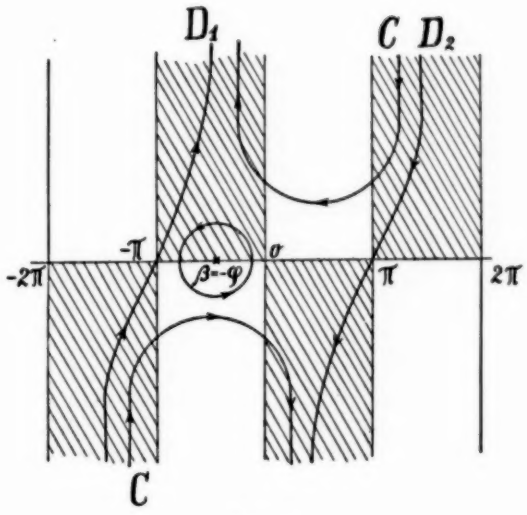

FIG. 2.

that (3) is really a solution of the problem. Firstly, at large distances from the diffracting body $\rho\to\infty$ the function $v(r,\psi-\psi_0)$ has to converge towards zero or towards the incident wave function u_0 according to whether the point under consideration is in the "illuminated" domain or in the "shadow" of geometrical optics. Secondly, the deviation of $v(r,\psi-\psi_0)$ from the terms of geometrical optics must contain for large ρ only outgoing and no incoming cylindrical waves ("emission-condition"). (If these conditions are fulfilled for the incident wave $v(r,\psi-\psi_0)$ they are automatically also fulfilled for the reflected wave $v(r,\psi+\psi_0)$ as a consequence of the boundary conditions on the surface of the wedge.) At last one has to postulate, that for $\rho\to 0$ v remains finite and $\rho\,\partial v/\partial\rho\to 0$.

Sommerfeld was able to fulfill all these conditions by making the following decisive Ansatz

$$v(\rho,\varphi)=\frac{1}{2\pi n}\int_C \frac{e^{i\rho\cos\beta}}{1-e^{-i(\beta+\varphi)/n}}d\beta. \tag{I}$$

Here C means the path in the plane of the complex variable. As indicated in the following Fig. 2, it consists of two parts. The shaded domains of the figure are those, where $\exp(i\rho\cos\beta)\to 0$ for $\rho\to\infty$.

We do not wish to repeat here the complete proof that $v(\rho,\varphi)$ fulfills all necessary conditions; instead we shall discuss only the behavior of this function for large values of ρ. To this

purpose it is important to realize that the integrand has poles at the places

$$\beta=-\varphi+2\pi nN;\quad (N=0,\pm1,\pm2,\cdots)$$

and that the behavior of v for large ρ is quite different, according to whether one of these poles falls into the interval $(-\pi, +\pi)$ or not. This fact becomes obvious, if we substitute for the path C of the integration the path D_1+D_2 of Fig. 2 and then add the residue of the pole, in case such a one is situated in the interval $(-\pi, \pi)$. In this way one gets

$$v=v^*+v_B \tag{8}$$

with

$$v^*=\begin{cases}\exp\left[i\rho\cos\left(\varphi+2\pi nN\right)\right], \\ \qquad\qquad \text{if}\quad -\pi<\varphi+2\pi nN<+\pi, \\ 0 \qquad\qquad \text{otherwise}\end{cases} \tag{9}$$

$$v_B=\frac{1}{2\pi n}\int_{D_1+D_2}\frac{e^{i\rho\cos\beta}}{1-e^{-i(\beta+\varphi)/n}}d\beta. \tag{10}$$

As a consequence of $n>1$, the inequality $-\pi<\varphi+2\pi nN<\pi$ can be fulfilled by not more than one single integer value of N, and it is easily seen that it is fulfilled for intervals of the length 2π and not fulfilled for intervals of the length $2\pi(n-1)$, which follow each other alternately. The former intervals we can call the illuminated ones, whereas the intervals of the latter kind indicate the shadow. Using this notation one has:

light for $\quad -\pi+2\pi nN<\varphi<\pi+2\pi nN,$ (11a)

shadow for

$$+\pi+2\pi nN<\varphi<\pi+2\pi nN+2\pi(n-1). \tag{11b}$$

The places $\varphi=-\pi+2\pi nN$ and $\varphi=+\pi+2\pi nN$ correspond to the *boundaries of the shadow*, with which we shall occupy ourselves particularly. A simple geometrical consideration shows indeed that $v^*(\rho, \psi-\psi_0)$ and $v^*(\rho, \psi+\psi_0)$ correspond in the physical angle interval exactly to the incident and to one or two reflected waves, respectively, as one expects them to occur according to geometrical optics.[2]

[2] As an example we point to the case $\pi-\theta<\psi_0<\pi$ where no boundary of the shadow of the incident wave, but two boundaries of the shadow of the reflected wave are situated in the physical angle interval $0<\psi<2\pi-\theta$.

For the following it is important that the functions v^* and v_B on the boundaries of the shadow $\varphi=\pm\pi+2\pi nN$ are both discontinuous, whereas their sum, as one recognizes from the path C in Fig. 1, is completely regular.

One can transform the expression (10) for the diffraction wave v_B by the substitution $\beta=\eta_1-\pi$ for D_1 and $\beta=\eta+\pi$ for D_2 and by adding the resulting integrands. In this way one obtains

$$v_B=\frac{1}{2\pi i}\cdot\frac{1}{n}\sin\frac{\pi}{n}\int_{-i\infty-\epsilon}^{+i\infty+\epsilon}\times\frac{e^{-i\rho\cos\eta}}{\cos\left(\pi/n\right)-\cos\left[(\eta+\varphi)/n\right]}d\eta, \tag{12}$$

where ϵ is an angle between zero and π. Physically interesting is the case of large values of ρ. The application of the method of steepest descent to the integral (12) gives in this case at once as contribution of the saddle-point $\eta=0$ the following asymptotic formula for v_B:

$$v_B=(2\pi\rho)^{-\frac{1}{2}}e^{-i(\rho+\pi/4)}\frac{n^{-1}\sin\pi/n}{\cos\pi/n-\cos\varphi/n}. \tag{13}$$

One can interpret this result as a cylindrical wave going out from the edge of the diffracting wedge with a characteristic dependence of the amplitude from the azimuth φ. The result is correct under the supposition

$$\rho(\cos\pi/n-\cos\varphi/n)^2\gg1,$$

it fails therefore in the neighborhood of the boundary of the shadow, where

$$\cos\pi/n=\cos\varphi/n.$$

Hence the problem arises to obtain such an asymptotic formula for the diffraction wave v_B for large ρ, which will be valid also in the vicinity of the boundary of the shadow.

This problem was attacked already by Sommerfeld[3] who tried to get an improved expression for v_B by a modification of the path of integration in (12). His results, however, were not satisfactory, because the neglected terms were not sufficiently small in comparison to the retained terms. We propose here a different treatment, which is based on a transformation of the

[3] A. Sommerfeld, Crelles J. **158**, 199 (1928).

integrand without changing the path of the integration.

We get an indication of this method by a previous result of Sommerfeld concerning the particular case of a half-plane ($\theta=0$, $n=2$). In this case he was able to transform the expression (12) into

$$v_B=-e^{i\pi/4}(2/\pi a)^{\frac{1}{2}}e^{i\rho\cos\varphi}\cdot\cos\varphi/2\cdot\int_{(a\rho)^{\frac{1}{2}}}^{\infty}e^{-i\tau^2}d\tau \quad (14)$$

with

$$a=1+\cos\varphi, \quad (15)$$

where the square root $a^{\frac{1}{2}}$ is always to be taken as positive. For the sum v^*+v_B therefore the expression results:

$$v(\rho,\varphi)=e^{i\pi/4}\pi^{-\frac{1}{2}}e^{i\rho\cos\varphi}\int_{-\infty}^{(2\rho)^{\frac{1}{2}}\cos\varphi/2}e^{-i\tau^2}d\tau, \quad (14a)$$

which is also regular for $\varphi=\pi$.

In the following section we shall define a system of functions $S_m(w)$ of the argument $w=\rho a$ in such a way that

$$S_0(w)=e^{iw}\cdot 2\int_{w^{\frac{1}{2}}}^{\infty}e^{-i\tau^2}d\tau. \quad (16)$$

$S_m(w)$ is for large w of the order of magnitude $w^{-\frac{1}{2}}$ whereas for small w, if $m>0$, $S_m(w)$ is of the order $w^{\frac{1}{2}}$. In Section 3 we shall finally establish for v_B a semi-convergent development, which holds also near the boundary of the shadow $\varphi=\pi$; namely,

$$v_B=e^{-i(\rho-\pi/4)}(2a)^{-\frac{1}{2}}\times\sum_{m=0}^{M}C_m(\varphi)S_m(\rho a)\rho^{-m}+R_M(\rho,\varphi). \quad (17)$$

The functions $C_m(\varphi)$ are regular for $\varphi=\pi$, but infinite for the other boundaries of the shadow $\varphi=\pi+2\pi nN$; $(N\neq 0)$. For every boundary of the shadow $\varphi=\pi+2\pi nN$ one can construct a development of the form (16); one has only to replace φ by $\varphi-2\pi nN$, with a given integer N, in the definition of a and in $C_m(\varphi)$. But in the general case there exists no development of this type which holds for *all* boundaries of the shadow simultaneously as well as for large ρ. This seems only to be possible for $n=2$, in which case the series (16) breaks off with the first term and reduces to (14).

2. The Functions $S_m(w)$

We shall develop Sommerfeld's solution into a series proceeding according to the system of functions $S_m(w)$, which is directly defined as a generalization of Fresnel's integral by

$$S_m(w)=e^{iw}w^m\int_w^{\infty}e^{-it}t^{-(m+\frac{1}{2})}dt=e^{iw}w^m\cdot 2\int_{w^{\frac{1}{2}}}^{\infty}e^{-i\tau^2}\tau^{-2m}d\tau. \quad (18)$$

In the physical application the variable w has a positive value and also the square root $w^{\frac{1}{2}}$ on the lower limit of the second integral is to be defined as positive. We shall prove that these functions have a simple connection with the confluent hypergeometric function $F(\alpha,\beta,x)$.[4] The latter can be defined as the solution of the differential equation

$$xF''+(\beta-x)F'-\alpha F=0, \quad (19)$$

which is given by the integral

$$F(\alpha,\beta,\gamma)=(2\pi i)^{-1}\Gamma(\beta)\int_C e^t t^{\alpha-\beta}(t-x)^{-\alpha}dt. \quad (20)$$

The powers of t are unambiguously defined by the statement that arg t and arg $(t-x)$ are running from $-\pi$ to $+\pi$ during the circulation around the point $t=0$ or $t=x$ along the path of integration. The path of integration can be chosen in different ways (Fig. 3). If it includes both singularities $t=x$ and $t=0$ (path C), one obtains a particular solution of (19), which can be represented also by the power series

$$F(\alpha,\beta,x)=1+\frac{1}{1!}\frac{\alpha}{\beta}x+\frac{1}{2!}\frac{\alpha(\alpha+1)}{\beta(\beta+1)}x^2+\cdots. \quad (21)$$

If, on the other hand, the path of the integration includes only one point (either the point $t=x$,

[4] Cf. Whittaker-Watson, *Modern Analysis*, 4th Edition (1927), Chap. 16; Frank-Mises, reference 1, Vol. 2, Chap. 26, 5. W. Gordon, Zeits. f. Physik **48**, 180 (1928). We omit on the left side of (20a) the factor $\frac{1}{2}$ in the definition of F_1 and F_2 used in the latter paper.

path C_1; or the point $t=0$, path C_2) whereas the other singularity is lying outside, one gets the solution F_1 or F_2, respectively:

$$(k=1,\ 2)\quad F_k(\alpha,\ \beta,\ x)=(2\pi i)^{-1}\Gamma(\beta)\times\int_{C_k}e^t t^{\alpha-\beta}(t-x)^{-\alpha}dt. \quad (20a)$$

As one sees in the figure, the relation holds:

$$F=F_1+F_2. \quad (22)$$

For large x there exist the, in general semi-convergent, developments:

$$F_1=\Gamma(\beta)/\Gamma(\alpha)\cdot e^z x^{\alpha-\beta}\times[1+(1-\alpha)(\beta-\alpha)/x+\cdots] \quad (23.1)$$

$$F_2=\Gamma(\beta)/\Gamma(\beta-\alpha)\cdot(-x)^{-\alpha}\times[1-\alpha(\alpha-\beta+1)/x+\cdots]. \quad (23.2)$$

The series for F_1 breaks off, if $\alpha=1,\ 2,\ \cdots$ or $\beta-\alpha=0,\ -1,\ -2\ \cdots$, whereas the series for F_2 breaks off for $\alpha=0,\ -1,\ -2\ \cdots$ and for $\beta-\alpha=1,\ 2,\ \cdots$.

We are interested here particularly in the case $\alpha=1$, $\beta=-m+\frac{3}{2}$, $x=iw$ with $m=0,\ 1,\ 2,\ \cdots$; that is to say in:

$$F_2(1,\ -m+\tfrac{3}{2},\ iw)=(2\pi i)^{-1}\Gamma(-m+\tfrac{3}{2})\times\int_{C_2}e^t t^{m-\frac{1}{2}}(t-iw)^{-1}dt. \quad (24)$$

In this case one can contract the path of integration to the negative real t axis and one obtains after the substitution $t\to -t$ and $t\to-\tau^2$ by taking care of $\Gamma(-m+\frac{3}{2})\Gamma(m-\frac{1}{2})=\pi(-1)^{m+1}$:

$$\begin{aligned}F_2(1,\ -m+\tfrac{3}{2},\ iw)&=\Gamma(m-\tfrac{1}{2})\int_0^\infty e^{-t}t^{m-\frac{1}{2}}(t+iw)^{-1}dt\\&=\Gamma(m-\tfrac{1}{2})\int_{-\infty}^{+\infty}e^{-\tau^2}\tau^{2m}(\tau^2+iw)^{-1}d\tau.\end{aligned} \quad (25)$$

On the other hand the path C_1 can in the case in question be replaced by a circle around the point $t=iw$ and gives (in accordance with the fact that the series (23.1) reduces here to the first term)

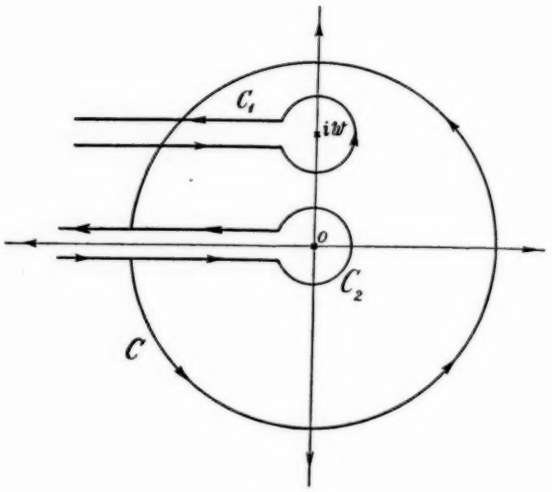

FIG. 3.

$$F_1(1,\ -m+\tfrac{3}{2},\ iw)=\Gamma(-m+\tfrac{3}{2})\times e^{i(m\pi/2-\pi/4)}w^{m-\frac{1}{2}}e^{iw}. \quad (26)$$

From (23.2) there follows the asymptotic expansion for large w

$$F_2(1,\ -m+\tfrac{3}{2},\ iw)=(m+\tfrac{1}{2})(iw)^{-1}\times[1-(m+\tfrac{1}{2})(iw)^{-1}+\cdots]. \quad (27)$$

Further one gets from (21) and (26) the development

$$\begin{aligned}F_2(1,\ -m+\tfrac{3}{2},\ iw)&=[1+(-m+\tfrac{3}{2})^{-1}iw\\&\quad+(-m+\tfrac{3}{2})^{-1}(-m+(5/2))^{-1}(iw)^2+\cdots]\\&\quad+\pi\Gamma(m-\tfrac{1}{2})^{-1}e^{-i(m\pi/2-\pi/4)}w^{m-\frac{1}{2}}e^{iw},\end{aligned} \quad (28)$$

which is especially practical for small values of w and shows that $F_2=1$ for $w=0$ when $m>0$, whereas for $m=0$ F_2 tends to infinity as $w^{-\frac{1}{2}}$ when w tends to zero.

The previously announced connection between the function $S_m(w)$, defined by (18), and the function $F_2(1,\ -m+\frac{3}{2},\ iw)$ defined by (25), is produced by the following transformation, which is a direct generalization of a transformation used by Sommerfeld in the case $m=0$. We substitute in (25) $t\to wt$ and get

$$F_2(1,\ -m+\tfrac{3}{2},\ iw)=\Gamma(m-\tfrac{1}{2})^{-1}w^{m-\frac{1}{2}}\times\int_0^\infty e^{-wt}t^{m-\frac{1}{2}}(t+i)^{-1}dt, \quad (25a)$$

hence

$$\frac{d}{dw}\left[e^{-iw}w^{-(m-\frac{1}{2})}F_2\right] = -\Gamma(m-\tfrac{1}{2})^{-1}e^{-iw}\int_0^\infty e^{-wt}t^{m-\frac{1}{2}}dt$$

$$= -(m-\tfrac{1}{2})e^{-iw}w^{-(m+\frac{1}{2})}.$$

By integration with respect to w we obtain from this fixing the constant of integration by means of the requirement that F_2 vanish for $w=\infty$:

$$F_2(1, -m+\tfrac{3}{2}, iw) = (m-\tfrac{1}{2})w^{m-\frac{1}{2}}e^{iw} \times\int_w^\infty e^{-it}t^{-(m+\frac{1}{2})}dt,$$

therefore according to (18)

$$S_m(w) = (m-\tfrac{1}{2})^{-1}w^{\frac{1}{2}}F_2(1, -m+\tfrac{3}{2}, iw). \quad \text{(II)}$$

From (27) and (28) there follows now immediately the behavior of $S_m(w)$ for small and for large values of w; in the latter case:

$$S_m(w) = -iw^{-\frac{1}{2}}[1-(m+\tfrac{1}{2})(iw)^{-1}+\cdots] \quad \text{for } w\gg 1, \quad \text{(27a)}$$

whereas in the former case one has

$$S_m(w)\sim(m-\tfrac{1}{2})^{-1}w^{\frac{1}{2}} \text{ for } w\sim 0 \text{ and } m>0. \quad \text{(28a)}$$

Finally from (25a) the useful recursion formula results:

$$\frac{d}{dw}[F_2(1, -m+\tfrac{3}{2}, iw)w^{-(m-\frac{1}{2})}] = -w^{-(m+\frac{1}{2})}(m-\tfrac{1}{2})F_2(1, -m+\tfrac{1}{2}, iw),$$

hence

$$\frac{d}{dw}[w^{-m}S_m(w)] = -(m+\tfrac{1}{2})w^{-(m+1)}S_{m+1}(w). \quad (29)$$

3. The Asymptotic Series

We start from the expression (12) for

$$v_B = \frac{1}{2\pi i}\cdot\frac{1}{n}\sin\frac{\pi}{n}\int_{-i\infty-\epsilon}^{+i\infty+\epsilon} \times\frac{e^{-i\rho\cos\eta}}{\cos(\pi/n)-\cos[(\eta+\varphi)/n]}d\eta. \quad (12)$$

The method of steepest descent, mentioned at the end of Section 1 consists of the introduction of the variable

$$s^2 = -i(1-\cos\eta), \quad s = e^{-i\pi/4}2^{\frac{1}{2}}\sin\eta/2, \quad (30)$$

so that

$$e^{-i\rho\cos\eta} = e^{-i\rho}e^{-\rho s^2}$$

$$\frac{d\eta}{ds} = e^{i\pi/4}2^{\frac{1}{2}}\left(1-\frac{i}{2}s^2\right)^{-\frac{1}{2}}. \quad \text{(30a)}$$

As path of integration, the real s axis from $-\infty$ to $+\infty$ may be chosen. The obvious procedure would be to develop the whole integrand, except the exponential function into powers of s. Corresponding to the poles of the integrand at $\eta=-\varphi \pm\pi+2\pi nN$ with $N=0, \pm1, \pm2\cdots$, the result obtained in this way would, however, become infinite for the values $\varphi=\pm\pi+2\pi nN$ which correspond to the boundaries of the shadow.

We can avoid the singularity of the result, at least at one given boundary of the shadow, by not developing the whole integrand (except the exponential function), but by also extracting explicitly the factor $(\cos\eta+\cos\varphi)^{-1}$ in such a way, that we put

$$f(s, \varphi) = \frac{\cos\eta+\cos\varphi}{\cos\pi/n-\cos[(\eta+\varphi)/n]}\cdot\frac{1}{\cos\eta/2} \quad (31.1)$$

and

$$v_B = \frac{1}{2\pi n}\sin\frac{\pi}{n}\cdot 2^{\frac{1}{2}}e^{-i(\rho-\pi/4)} \times\int_{-\infty}^{+\infty} e^{-\rho s^2}f(s, \varphi)\frac{ds}{s^2+ia} \quad (31.2)$$

with the abbreviation $a=1+\cos\varphi$ introduced in (15), and with $\cos\eta+\cos\varphi = a-is^2$.

The essential effect of the addition of the factor $\cos\eta+\cos\varphi$ in the definition of $f(s, \varphi)$ is the fact that this function is now regular for $\varphi=\pm\pi$ and $\eta=0$ and becomes only infinite for $\eta=0$, if $\varphi=\pi+2\pi nN$, with N integer, but Nn not integer (particularly $N\neq 0$). Therefore our result can be used for $\varphi=\pm\pi$ and fails only on the other boundaries of the shadow. If one replaces

in the definition of a and of $f(s, \varphi)$ and, in all following equations the variable φ by $\varphi - 2\pi n N_0$, one gets for every given integer N_0 a development of v_B, which holds at $\varphi = \pm\pi + 2\pi n N_0$ instead of at $\varphi = \pm\pi$.

If we consider now $f(s, \varphi)$ as developed in powers of s according to

$$f(s, \varphi) = \sum_{m=0}^{\infty} e^{im\pi/4} A_m(\varphi) s^m \tag{32}$$

and substitute this development in (31.2), the terms with odd m cancel when integrating and the terms with even m give after the substitution $s = \tau\rho^{-\frac{1}{2}}$ integrals of the type which occurred in (25). By using formula (II) of Section 2 one gets finally (compare (17)):

$$v_B = (2\pi a)^{-\frac{1}{2}} n^{-1} \sin\frac{\pi}{n} e^{-i(\rho - \pi/4)} \cdot \sum_m i^m \Gamma(m + \tfrac{1}{2}) \pi^{-\frac{1}{2}} A_{2m}(\varphi) S_m(\rho a) \rho^{-m}. \tag{33}$$

The circumstance, however, that the series (32) converges only on a part of the path of integration, has the consequence that the series (33) in general has the meaning of a semi-convergent development for large values of ρ.

We did not calculate the rather complicated expression for $A_{2m}(\varphi)$ for a general m; but we wish to emphasize that these functions have a finite value for $\varphi = \pm\pi$. One recognizes this most simply by putting $\varphi = \pm\pi$ on the left side of (32) and developing in powers of s afterwards. For the first two functions $A_0(\varphi)$ and $A_2(\varphi)$ one gets from (30) and (32)

$$A_0(\varphi) = \frac{1 + \cos\varphi}{\cos\pi/n - \cos\varphi/n}; \qquad (2a)^{-\frac{1}{2}} A_0(\varphi) = \frac{|\cos\varphi/2|}{\cos\pi/n - \cos\varphi/n} \tag{34.1}$$

$$A_2(\varphi) = -\frac{1}{4} A_0(\varphi) - \frac{1}{n^2} \cos\frac{\varphi}{n} \frac{1 + \cos\varphi}{(\cos\pi/n - \cos\varphi/n)^2} + \frac{2[(1/n)\sin\varphi/n]^2 (1 + \cos\varphi) - (\cos\pi/n - \cos\varphi/n)^2}{(\cos\pi/n - \cos\varphi/n)^3}, \tag{34.2}$$

hence

$$[(2a)^{-\frac{1}{2}} A_0(\varphi)]_{\varphi = \pi \mp \epsilon} = \mp\tfrac{1}{2}(1/n \cdot \sin\pi/n)^{-1} \tag{34.3}$$

$$A_2(\pi) = -\frac{1}{2} \frac{\cos\pi/n}{\sin^2\pi/n}. \tag{34.4}$$

The finiteness of the series for v_B at $\varphi = \pm\pi$ (where $a = 0$), follows from the fact that, when $m > 0$, the factor $a^{-\frac{1}{2}}$ is compensated by $S_m(\rho a)$ because of $S_m(\rho a) \sim (m - \frac{1}{2})^{-1} (\rho a)^{\frac{1}{2}}$ and when $m = 0$, is compensated by $A_0(\varphi)$.

In the particular case $n = 2$ one has

$$\tfrac{1}{2}[f(s) + f(-s)] = -2\cos\varphi/2$$

and therefore *all* $A_{2m}(\varphi)$ *vanish in this case identically for* $m > 0$ and the whole series (33) reduces to its first term.

Writing out explicitly the terms of (33) for $m = 0$ and $m = 1$ we get according to (16)

$$v_B = \pi^{-\frac{1}{2}} e^{i\pi/4} (1/n \sin\pi/n) \times \left\{ \frac{2|\cos\varphi/n|}{\cos\pi/n - \cos\varphi/n} e^{i\rho\cos\varphi} \int_{(a\rho)^{\frac{1}{2}}}^{\infty} e^{-i\tau^2} d\tau + \frac{i}{2}(2a)^{-\frac{1}{2}} A_2(\varphi) S_1(\rho a) \frac{e^{-i\rho}}{\rho} + \cdots \right\}. \tag{35}$$

For large values of ρa one gets from (27a)

$$v_B = (2\pi\rho)^{-\frac{1}{2}} e^{-i(\rho + \pi/4)} \times \frac{(1/n)\sin\pi/n}{\cos\pi/n - \cos\varphi/n} [1 + (*)\rho^{-1} + \cdots]$$

in accordance with (13). On the other hand one has for $\varphi = \pi$ because of

$$\int_0^{\infty} e^{-i\tau^2} d\tau = \tfrac{1}{2}\pi^{\frac{1}{2}} e^{-i\pi/4}:$$

$$v_B = \mp\tfrac{1}{2} e^{-i\rho} + (i/2)(2\rho)^{-\frac{1}{2}} \times e^{-i\rho}(1/n)\cot\pi/n + \cdots. \tag{36}$$

The upper (lower) sign holds when approaching $\varphi=\pi$ from the illuminated (shadowy) side. The discontinuity of v_B on the boundary of the shadow is just compensated by the discontinuity of v^* in such a way that the sum $v=v^*+v_B$ is continuous. Further we can confirm a general result of Sommerfeld, according to which on the boundary of the shadow itself for large ρ the relation asymptotically holds

$$v=\tfrac{1}{2}v^*(\rho\gg 1,\ \varphi=\pm\pi+2\pi nN). \qquad (37)$$

That not only for the function v_B, but also for all its derivatives the discontinuities on the boundary of the shadow are compensated for by those of v^*, can be confirmed by using Eq. (28). We do not want, however, to enter the details of this proof here.

Numerical estimates of $A_2(\varphi)$ gave the result that already the second term of the series (35) can be neglected in all practical cases so that the proposed task of finding a representation of the diffraction wave, which can be used for the transition from light to shadow and at large distances, can be considered as practically solved by the principal term of our series, as given in (35).

DECEMBER 1, 1938 PHYSICAL REVIEW VOLUME 54

On the Anomalous Propagation of Phase in the Focus

A. RUBINOWICZ
John Casimir University, Lwów, Poland

(Received October 11, 1938)

The problem of anomalous phase propagation of a spherical wave at the focus has been discussed for the case of a diffracting aperture of arbitrary shape. The solution given by Kirchhoff's integral has been split up into an "incident light wave," which shows the distribution of light to be expected according to geometrical optics and a "diffracted wave," which may be thought of as due to scattering of the incident wave at the diffracting edge. A sudden change of phase by π has been shown to occur already in the incident wave. Thus we may, in this sense, consider this phenomenon as a geometric optical one.

The case of a circular diffracting aperture, the focus lying on the normal through its center, which has been treated usually, appears to be not very suitable for an experimental investigation of the discussed phenomenon. It is this particular shape of the diffracting edge, which produces diffraction phenomena of considerable light intensity along the optical axis. These, however, are not because of the existence of a focus, but only because of the particular shape of the diffracting aperture.

I. INTRODUCTION

SOMMERFELD'S first great scientific achievement was the solution of the problem of diffraction at a perfectly conducting half-plane by the methods of exact analysis.[1] On the occasion of his jubilee it may thus be appropriate to present a note, which deals with a related problem and is based on a paper[2] I wrote while staying at his institute at Munich. The problem in question is the anomalous phase propagation of a spherical wave at the focus, which was discovered by Gouy in 1890. As appears from the very rich literature[3] on this subject, the phenomenon is completely describable by wave optics. Still, I do not believe that these papers satisfy as yet our want for a simple, plausible interpretation of the problem.[4] An attempt, therefore, will be made in this note to fill this gap by following a new method of approach, starting from Kirchhoff's theory of diffraction

[1] A. Sommerfeld, Math. Ann. **47**, 317 (1896).
[2] A. Rubinowicz, Ann. d. Physik **53**, 257 (1917) and **73**, 339 (1924), to be referred to as I and II in the text.
[3] For extensive literature see F. Reiche, Ann. d. Physik **29**, 65, 401 (1909) and J. Picht, *Optische Abbildung* (Braunschweig, 1931).
[4] An elementary suggestive representation by means of Fresnel's zones has, however, been given by A. D. Fokker, Physica **3**, 334 (1923); **4**, 166 (1924).

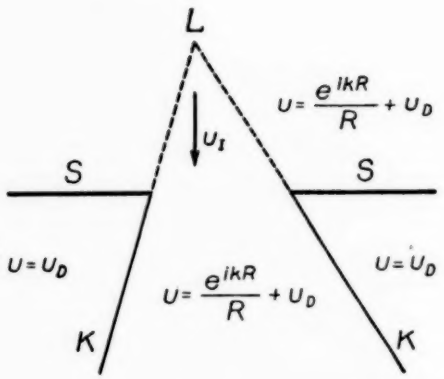

FIG. 1. Diffraction of a divergent spherical wave. L source of light, S diffracting screen, K boundary of the shadow. The arrow u_I shows the direction of propagation of the incident light wave.

and, besides, generalizing the treatment so as to apply to an arbitrary diffracting aperture.

We start from Kirchhoff's theory of diffraction but generalize the treatment so as to apply it to an arbitrary diffracting aperture.

At all those points where light should be present according to the laws of geometrical optics, the incident wave u_I is given by the solution of $\Delta u+k^2u=0$ for direct propagation of light, that is, if the time factor is e^{-ivt}, e^{ikR}/R for a point source; R is the distance from the source. It is supposed to vanish everywhere else. The wave function u_I will thus decrease discontinuously by e^{ikR}/R while passing from the "light cone" across the "boundary of the shadow" into the "shadow" (Fig. 1).

Concerning the shape of the geometrical boundary of the shadow we note that it is given by the surface of the cone K, which is generated by the lines joining the source of light L and the points of the diffracting edge D. The part of the cone between L and D has to be regarded as cut off.

The diffraction wave u_D can be considered as due to scattering of the incident light wave u_I at the different points of the diffracting edge D; it is composed of the secondary elementary waves produced there. Since we expect $u=u_I+u_D$ taken as a whole to be continuous on the boundary of the shadow, u_D must be discontinuous in such a way as to compensate the discontinuity of u_I. It ought thus to increase suddenly by e^{ikR}/R at the passage from the light cone into the shadow.

Let us suppose that the diffraction of a *convergent* spherical wave may be treated just like that of a *divergent* spherical wave considered above by representing the corresponding wave function u as a sum of u_I and u_D. Before determining these functions we have again to ascertain the boundary of the geometrical shadow. Obviously this will be formed by the double cone K_1+K_2, generated, as above, by joining its vertex, now the focus F, with D (Fig. 2), which simultaneously cuts off the remaining part of the cone K_1.

Let us represent the incident light wave by $u_{I_1}=e^{-ikR}/R$ (R now denoting the distance from the focus F) in that portion of space which is bounded by the cone K_1 and the screen S and contains also the interior of the cone K_1. The minus sign had to be put into the exponent because the wave is now convergent; the time factor e^{ivt} is the same as above. When passing through the cone K_1 into the geometrical shadow u_{I_1} is again supposed to disappear.

Let us now further assume that a diffracted wave u_D is produced by scattering of the incident wave u_{I_1} at the diffracting edge D. If $u_{I_1}+u_D$ is to be continuous at the boundary of the shadow, u_D ought to increase suddenly by e^{-ikR}/R when passing through the cone K_1 into the geometrical shadow. Now let us still make the plausible assumption that the elementary wavelets scattered at the different points of the diffracting edge, and with them also the diffracted wave u_D consti-

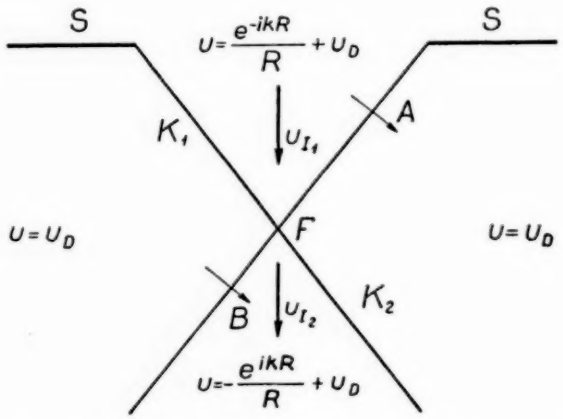

FIG. 2. Diffraction of a convergent spherical wave. F focus of the incident spherical wave, S diffracting screen, K_1 and K_2 boundaries of the shadow. The two arrows u_{I_1} and u_{I_2} show the direction of propagation of the incident light wave in the cones K_1 and K_2.

tuted by them, do not show any peculiar irregularity in their behavior while passing through the focus F. Then the discontinuity of u_D ought to be continued regularly on K_2. Since R is always positive according to our notation, a sudden increase of u_D by e^{-ikR}/R on the cone K_1 will correspond to that by e^{+ikR}/R on K_2. The direction of the passages through the surfaces is that indicated by the arrows A and B in Fig. 2. Otherwise the diffracted wave would be discontinuous at the focus F and travel along the cone K_2 towards the focus, that is, in a direction opposite to that in which it should go.

And now to our decisive remark: Outside the double cone K_1+K_2, that is, within the geometrical shadow, we have $u=u_D$, since for a passage through K_1 out of the light cone, like that shown by the arrow A, the discontinuity of u_{I_1} is just compensated by that of u_D. Now, u_D increases discontinuously at the passage through K_2 indicated by arrow B. To compensate this discontinuity on K_2 we shall have to put $u=u_D-e^{+ikR}/R$ inside K_2. Hence we see that the incident light must be represented by $u_{I_2}=-e^{+ikR}/R$ inside the cone K_2. The minus sign of u_{I_2} indicates that u_{I_1} and u_{I_2} differ in phase by π. We may thus summarize our results in the statement: The incident light wave is represented by a convergent spherical wave inside K_1 and by a divergent spherical wave inside K_2, which differ in phase by half a period.

Since, in general, the diffraction wave vanishes in the limiting case of infinitely small wavelengths, i.e., for $k\to\infty$, while the incident wave still persists, the phase anomaly will have to be considered as a phenomenon, which has to be taken into account in the geometric optical treatment of the problem as a sudden change in phase by half a period. It must, however, be said that this statement cannot apply at all those points of space, where, independently of the value of k, the intensity of the diffracted wave is of the same order of magnitude as that of the incident light wave. This is only possible at points, where the secondary waves from at least some finite portion of D arrive with all their phases equal. In general the focus F will be such a point, but in some particular cases still other points of this kind exist. In the case exclusively considered hitherto, for instance, where the diffracting aperture is circular and the focus lies on the normal through its center, all the points on the optical axis thus formed will be of that kind.

In the following paragraphs a mathematical formulation and some extension of the propositions made will be given.

II. Incident and Diffracted Wave

The distribution of light, the case of diffraction of a convergent spherical wave, at those points of space which lie in front of the screen S and the surface, extended over the diffracting edge D, say f, is given by Kirchhoff's integral:

$$u_K=\frac{1}{4\pi}\int_f\left(\frac{e^{-ik\rho}}{\rho}\frac{\partial}{\partial n}\frac{e^{ikr}}{r}-\frac{e^{ikr}}{r}\frac{\partial}{\partial n}\frac{e^{-ik\rho}}{\rho}\right)df, \quad (1)$$

where ρ is the distance of the surface element df from the focus F, r the distance of df, from the point P, at which we calculate u_K and n the normal of f in the direction away from the focus F. No restricting assumptions will be made at present concerning the shape of the diffracting edge D.

To transform this expression we shall first represent the incident light wave, say u_I, by means of surface integrals. Making use of the same notation as in Section I, we put

$$u_I=u_{I_1}+u_{I_2}, \quad (2a)$$

where u_{I_1} and u_{I_2} are defined as follows:

$$\begin{aligned} u_{I_1}&=e^{-ikR}/R \text{ inside the region bounded by } K_1 \text{ and } f, \\ &=0 \text{ at all other points.} \end{aligned} \quad (2b)$$

$$\begin{aligned} u_{I_2}&=-e^{+ikR}/R \text{ inside the cone } K_2, \\ &=0 \text{ at all other points.} \end{aligned} \quad (2c)$$

To obtain integral representations for (2b) and (2c) we make use of the following well-known theorem: Let u^* be a regular solution of $\Delta u+k^2u=0$ inside the region enclosed by the surface Φ. Then the integral

$$\frac{1}{4\pi}\int_\Phi\left(u^*\frac{\partial}{\partial n}\frac{e^{ikr}}{r}-\frac{e^{ikr}}{r}\frac{\partial u^*}{\partial n}\right)df \quad (3)$$

represents the function u^* at the point P (whose

distance from df is denoted by r) if P lies inside Φ, and vanishes for P outside Φ. n is the external normal of Φ.

Using (3) we get immediately·

$$u_{I_1}=\frac{1}{4\pi}\int_{f+K_1+\Omega_1}\left(\frac{e^{-ik\rho}}{\rho}\frac{\partial}{\partial n}\frac{e^{ikr}}{r}-\frac{e^{ikr}}{r}\frac{\partial}{\partial n}\frac{e^{-ik\rho}}{\rho}\right)df, \quad (4a)$$

$$u_{I_2}=-\frac{1}{4\pi}\int_{K_2+\Omega_2}\left(\frac{e^{ik\rho}}{\rho}\frac{\partial}{\partial n}\frac{e^{ikr}}{r}-\frac{e^{ikr}}{r}\frac{\partial}{\partial n}\frac{e^{ik\rho}}{\rho}\right)df, \quad (4b)$$

where Ω_1 and Ω_2 are spherical surfaces round F as their center. They had to be introduced on account of the singularity of $e^{\pm ik\rho}/\rho$ at F, which is thus cut off.

Concerning the integral representations (4a, b) we have to make the following remarks:

First the integral over f in (4a) is identical with u_K.

Secondly the integrals over K_1 and K_2 in (4a, b), which we shall denote by $-u_{D_1}$ and $-u_{D_2}$ yield the expressions

$$u_{D_1}=-\frac{1}{4\pi}\int_{K_1}\frac{e^{ik(r-\rho)}}{\rho}\left(\frac{ik}{r}-\frac{1}{r^2}\right)\cos(n,r)df, \quad (5a)$$

$$u_{D_2}=\frac{1}{4\pi}\int_{K_2}\frac{e^{ik(r+\rho)}}{\rho}\left(\frac{ik}{r}-\frac{1}{r^2}\right)\cos(n,r)df, \quad (5b)$$

since $(\partial/\partial n)(e^{\pm ik\rho}/\rho)$ vanishes on K_1 and K_2 and further

$$\frac{\partial}{\partial n}\frac{e^{ikr}}{r}=e^{ikr}\left(\frac{ik}{r}-\frac{1}{r^2}\right)\cos(n,r).$$

n is here the external normal of K_1 or K_2. Its direction is reversed as we pass along a generator of our cone through the focus F.

Finally we note that the two integrals over Ω_1 and Ω_2 yield $-$ or $+$ $\omega e^{ikR}/4\pi R$, where ω is the solid angle subtended by these surfaces.

Now, on the basis of these remarks and Eq. (2a) we obtain adding (4a) and (4b):

$$u_K=u_I+u_{D_1}+u_{D_2},$$

where the sum $u_{D_1}+u_{D_2}$ can be reduced to a line integral over the diffraction edge D. To perform this transformation we shall introduce on our double cone K_1+K_2, an orthogonal coordinate system, given by the distance ξ from the focus F and the lines of intersection σ of $\xi=$ const. with K_1 and K_2. The sign of ξ may be chosen so that $\xi=-\rho$ on K_1 and $\xi=+\rho$ on K_2.

To obtain a suitable expression for df we consider the area subtended on our cone K_1+K_2 by an element ds of the diffracting edge D. Let us denote by ρ_s the distance of ds from the focus F. Then the line element of $d\sigma$ (Fig. 3) on our area is given by $d\sigma=(\xi/\rho_s)\sin(\rho_s, ds)\,ds$ and therefore $df=d\xi d\sigma$ by $df=(\xi/\rho_s)\sin(\rho_s, ds)d\xi ds$. The sign of df represented in this way will be positive on K_2 and negative on K_1, while it is always positive in the integrals (5a, b) for u_{D_1} and u_{D_2}. But in these integrals also $\cos(n, r)$ changes its sign at the focus F, where, as we have said, the direction of n is suddenly reversed. Choosing now that direction of n, which coincides with its former direction on K_1, opposite to it on K_2, we shall have to put, with these new assumptions concerning n

$$-\cos(n,r)(\xi/\rho_s\sin(\rho_s, ds)ds d\xi$$

instead of $\cos(n, r)df$ in our integrals. Considering further that now

$$\cos(n,r)=(r_s/r)\cos(n,r_s),$$

where r_s is the distance of the point P from ds, we obtain the expression

$$u_D=-\frac{1}{4\pi}\int_D ds\sin(\rho_s, ds)\cos(n,r_s)\times\frac{r_s}{\rho_s}\int_{-\rho_s}^{+\infty}d\xi e^{ik(\xi+r)}\left(\frac{ik}{r^2}-\frac{1}{r^3}\right)$$

for the sum $u_D-u_{D_1}+u_{D_2}$. Since now

$$r^2=r_s^2+(\xi+\rho_s)^2-2r_s(\rho_s+\xi)\cos(r_s,\rho_s)$$

we get finally, on following the method given in I (p. 261)

$$u_D=\frac{1}{4\pi}\int_D\frac{e^{-ik\rho_s}}{\rho_s}\frac{e^{ikr_s}}{r_s}\frac{\cos(n,r_s)}{1-\cos(r_s,\rho_s)}\times\sin(\rho_s, ds)ds. \quad (6)$$

Kirchhoff's diffraction integral may thus be given the form

$$u_K=u_I+u_D, \quad (7)$$

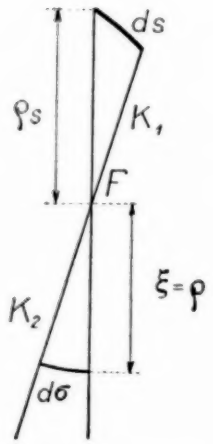

FIG. 3. To the equation $d\sigma=(\xi/\rho_s)\sin(\rho_s, ds)$.

where u_I represents a wave of the kind to be expected according to geometrical optics and u_D a wave, which may be thought of as due to scattering of the incident wave by the diffracting edge. The secondary waves, which originate at the different elements ds of the edge, are seen to be asymmetric on account of the direction factor

$$\frac{\cos(n, r_s)}{1-\cos(r_s, \rho_s)}. \tag{8}$$

This becomes infinite on that generator of the double cone K_1+K_2, which belongs to the element ds, since there $\cos(r_s, \rho_s)=1$.

III. Approximate Representation of the Diffraction Wave

If we want to discuss the distribution of light for the particular case of diffraction of a convergent spherical wave, we have to resort to appropriate approximate formulae for the diffraction wave u_D (or for the complete wave function u_K). For all those points of space, at which the method of stationary phase can be applied to the integral (6) either directly or indirectly, such formulae are readily obtained in the same way as for a divergent spherical wave (cf. II). The necessary condition for this is that the phase factor $k(r_s-\rho_s)$ varied rapidly enough as we pass along the diffracting edge D. Then only the immediate vicinity of those points P_ν, where the relatively slowest variation of phase occurs, that is, where

$$d(r_s-\rho_s)/ds=0 \tag{9}$$

or

$$\cos(r_s, ds)=\cos(\rho_s, ds), \tag{9a}$$

will yield a noticeable contribution to the integral (6).

Hence it appears that the light scattered by a particular element ds of the edge is spread over a circular cone, whose vertex lies in that element. There is also the line joining the element ds and the focus F among the generators of this cone. Its direction coincides with that of the light ray of the convergent spherical wave, incident upon ds. It is in such half-cones that light is reflected in the points of a reflecting curve, e.g., a wire, according to the concepts of geometrical optics. We may thus call this half-cone the reflection cone.

If we want to apply directly the method of stationary phase to the integral (6), we have to assume that apart from the factor $\exp[ik(r_s-\rho_s)]$ the integrand varies so slowly that it may be taken out of the integral. This assumption is, however, not fulfilled in the points near the boundary of the shadow, where the direction factor (8) becomes infinite. Therefore we cannot apply this method directly at those very points in which we are most interested, at the treatment of diffraction problems. Bur fortunately there is an indirect way of deriving approximate formulae, which can be applied even at points in the boundary of the shadow itself. By following the method applied in II, (p. 352) we obtain for the contribution $u_D|_\nu$ of the effective region round the point P_ν to u_D the expression

$$u_D|_\nu=\frac{1}{2r_s\rho_s(\zeta'')^{\frac{1}{2}}}\frac{(R+r_s-\rho_s)^{\frac{1}{2}}}{1-\cos(r_s, \rho_s)}\cos(n, r_s)$$
$$\times\sin(\rho_s, ds)e^{i(-kR+3\pi/4)}$$
$$\times\int_{+\infty}^{+[(2k/\pi)(R+r_s-\rho_s)]^{\frac{1}{2}}} e^{i(\pi/2)v^2}dv, \tag{10}$$

if the points considered lie inside the cone K_1.[5]

[5] The frequent appearance of Fresnel's integrals in Kirchhoff's theory of diffraction is due to the fact that they occur in all those cases where the incident light is reflected by the diffracting edge.

Here we have put

$$\zeta''=\frac{d^2}{ds^2}(r_s-\rho_s)=\sin^2(\rho_s, ds)\left(\frac{1}{r_s}-\frac{1}{\rho_s}\right)+\frac{1}{H}(\cos(r_s, H)-\cos(\rho_s, H)), \quad (11)$$

where H is the radius of curvature of the diffracting edge in the point P, taken as a vector in the direction of the principal normal at P_ν.

If we want to calculate u_D by means of (10) at some point of space, we have first to ascertain from which points P_ν of the diffracting edge the incident light is reflected in cones, to the point of space considered. u_D will then be given as the sum of the respective $u_D|_\nu$.[6] For a circular diffracting aperture, the focus F lying on the normal through its center, formula (10) may be found in a paper by Schwarzschild.[7] Then only two points P_ν need be considered and u_K consists simply of the incident wave and the waves, reflected at these two points.

For the derivation of (10) it has been supposed that in the vicinity of P_ν, the phase may be represented with sufficient accuracy by

$$r_s-\rho_s=(r_s-\rho_s)_{P_\nu}+\frac{1}{2}\frac{d^2}{ds^2}(r_s-\rho_s)(s-s_\nu)^2=(r_s-\rho_s)_{P_\nu}+\tfrac{1}{2}\zeta''(s-s_\nu)^2$$

in which use is made of (9). This approximation fails, however, for ζ'' vanishes on the reflection cone, as appears already, from the occurrence of $(\zeta'')^{\frac{1}{2}}$ in the denominator of (10). The points $\zeta''=0$ are distinguished as forming the intersection lines of two "infinitely neighboring" reflection cones. The images of the incident convergent light wave at its reflection from a definite element ds of the diffracting edge lie on these lines, which may thus be referred to as focal lines. They separate the reflection cone into two domains, one of positive and one of negative ζ''. It has, however, been supposed that $\zeta''>0$ at the derivation of (10). Now, on investigating the case $\zeta''<0$, we find that there is a change of phase by $\pi/2$ occurring at the passage across the focal line $\zeta''=0$. The phase anomaly is thus shown, not only by the incident, but also, eventually, so to say on a small scale, by the diffracted wave.[8]

The method of stationary phase fails near the focus F. In the focus itself, where $r_s=\rho_s$, all the wavelets originating at different points of the diffracting edge arrive with equal phase. Also near the focus their phases will differ little, so that it will be necessary to consider the diffracting edge as a whole and not only some small effective regions. A general representation like that possible at those points of space, where the method of stationary phase applies, does no longer seem to be feasible. Besides, it will be advisable to use u_K and not u_D near F, where the latter becomes infinite according to (2) and (7).

We have still to say a few words about the special case of a circular diffracting aperture, the focus of the convergent spherical wave lying on the normal through the center of the circle. For every point on the optical axis thus formed not only the phase $r_s-\rho_s$, but the whole integrand in (6) will be constant. Then the amplitude of u_D does no longer depend upon the wave-length and will be comparable with that of the incident light. These peculiar conditions on the optical axis are due, however, to the particular shape of the diffracting edge and not to the existence of a focus. This large amplitude of u_D is also responsible for the fact that points on the axis are favored at all observations. The large focal length and the smallness of the aperture used at most observations will add to this peculiarity, since the phase variations of the elementary wavelets for a point at a given distance from the axis decrease with the distance from and the smallness of the diffracting aperture. These two circumstances favor an extension of the conditions on the axis into the neighboring portions of space.

[6] (10) enables us to treat approximately diffraction—though only if it is an extrafocal one—by more than one diaphragm, a problem rather difficult to solve by Fresnel-Kirchhoff's original methods.

[7] K. Schwarzschild, Sitzungsber. d. math. phys. Klasse d. k. bayerischen Akad. d. Wiss. **28**, 271 (1898).

[8] This anomaly of phase on a small scale occurs, of course, also at the diffraction of a divergent spherical wave, which has been treated in II.

A Simple Derivation of the Maxwell-Boltzmann Law

E. U. CONDON
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received August 11, 1938)

The Maxwell-Boltzmann law is derived in a direct and simple way from the usual postulates that the mechanical system has a discrete system of allowed states and that each of these states has equal *a priori* weight in calculation of statistical averages.

THE Maxwell-Boltzmann distribution law plays such a fundamental role in the entire theory of thermal phenomena that it is important to have as simple an approach to it as possible. In this paper, a method of derivation is presented which is believed to be essentially new[1] and which shows in an extremely simple manner the use of the assumptions involved.

There are only two basic assumptions needed: (1) The dynamical system in question is governed by quantum mechanics and, being a closed system, has, therefore, a discrete spectrum of allowed energy levels, (2) in calculating statistical averages, each state corresponding to one linearly independent wave function orthogonal to all the others, that is, consistent with known features of the problem, is to be given the same weight.

Let us first consider the case of a large number N of noninteracting molecules in a cubical box of edge L and volume $V=L^3$. The wave function for an allowed state of a single molecule is

$$u_{lmn}(x, y, z)=\left(\frac{2}{V}\right)^{\frac{1}{2}} \sin\frac{l\pi x}{L} \sin\frac{m\pi y}{L} \sin\frac{n\pi z}{L} \quad (1)$$

and the corresponding energy is

$$W(l, m, n)=(h^2/8\mu L^2)(l^2+m^2+n^2). \quad (2)$$

The allowed states for N noninteracting molecules in a box will have as wave functions a continued product of such one-particle wave functions (1) and the total energy of the N particles will be a sum of such one-particle energy expressions (2). Neglecting the symmetry restrictions on the wave function which would give rise to Einstein-Bose and Fermi-Dirac statistics, there will be one state of the system for each possible complete set of quantum numbers consisting of N individual sets $(l_\alpha, m_\alpha, n_\alpha)$, where $\alpha=1, 2, \cdots, N$ and the $3N$ quantum numbers $l_\alpha, m_\alpha, n_\alpha$ range independently over all positive integral values.

It is important to know the total number of states of the system whose total energy is equal to or less than W for values of W large compared to the interval between energy levels. This is found by the method that has often been used in statistical mechanics. We introduce a $3N$ dimensional space. In this space the state of the system whose quantum numbers are $l_1m_1n_1$, $l_2m_2n_2$, $\cdots$, $l_Nm_Nn_N$ is associated with the point having this set of positive integers for Cartesian coordinates. There is thus one state per unit volume in this space. The number of states for which the energy is less than W is thus equal to 2^{-3N} times the volume of a $3N$ dimensional sphere of radius $(8\mu L^2W/h^2)^{\frac{1}{2}}$.

Denoting the number of states for the N particle system whose energy is less than W by $C_N(W)$, it is evident from dimensional considerations alone that $C_N(W)$ is proportional to $W^{3N/2}$ which fact is all that is needed in the derivation of the distribution law to be given below. From geometry it is known that the volume of an n dimensional sphere of radius r is

$$(\pi^{n/2}r^n)/\Gamma\left(\frac{n}{2}+1\right) \quad (3)$$

[1] *Note added in proof:*—Since this was written, I have noticed that a similar derivation is given by Kennard in his new book *Kinetic Theory of Gases*, p. 390. I have also had an interesting discussion with Professor G. E. Uhlenbeck, who remarks that the approach followed here is essentially a modern version of part of Maxwell's work as presented, for example, in Jeans' *Dynamical Theory of Gases*, Chapter V, p. 119 *et seq.* He also remarked on its connection with the discussion of statistical distribution of energy among a small number of particles as worked out for a problem in nuclear physics in a paper by Uhlenbeck and Goudsmit, pp. 201–211 in the *Zeeman Verhandelingen* (Martinus Nijhoff, The Hague, 1935).

and, therefore, the exact formula for $C_N(W)$ is

$$C_N(W)=\frac{1}{\Gamma\left(\frac{3N}{2}+1\right)}\left[\frac{(2\pi\mu W)^{\frac{1}{2}}L}{h}\right]^{3N}. \quad (4)$$

Now let us consider a system consisting of $(N+1)$ noninteracting molecules in the box and ask for the probability that the extra molecule have an energy between w and $w+dw$ when all we know of the system is that the total energy is W. The probability, according to the statistical postulate, will be proportional to the number of states of the composite system for which w lies between w and $w+dw$ and hence the energy w_N of the other N molecules lies between $W-w$ and $W-w-dw$. Therefore, writing $P(w)dw$ for the probability that the energy of the extra molecule lie in this range, we have

$$P(w)dw\sim C_N'(W-w)C_1'(w)dw, \quad (5)$$

where $C_N'(w)$ is the derivative of $C_N(w)$ with respect to w. Using only that part of (4) which makes $C_N(W)$ proportional to $W^{3N/2}$, we have

$$P(w)dw\sim(1-w/W)^{3N/2-1}w^{\frac{1}{2}}dw. \quad (6)$$

Let us define an amount of energy called kT by the equation

$$W=3NkT/2. \quad (7)$$

Then, since N is very large, we can recognize that the first factor is equal to the exponential factor of the Boltzmann distribution, so

$$P(w)dw\sim e^{-w/kT}w^{\frac{1}{2}}dw, \quad (8)$$

which is the usual Maxwellian distribution of velocities.

This mode of derivation brings out clearly that the $w^{\frac{1}{2}}$ factor arises from the fact that there are more states of the single molecule available in unit energy range at higher energies than at lower energies, whereas the exponential factor arises from the fact that if the single molecule gets more energy, there is necessarily less left for the other N molecules, and thus they are required to be in a range of energy where fewer states per unit energy range are available for them.

It is now a simple calculation to normalize (8) and find that

$$P(w)dw=(2/\pi^{\frac{1}{2}})e^{-w/kT}(w/kT)^{\frac{1}{2}}d(w/kT) \quad (9)$$

and to calculate that the mean energy of a single molecule is

$$w_{Av}=3kT/2. \quad (10)$$

Physically, the N molecules in the composite system can be regarded as the perfect gas thermometer with which the single molecule is in thermal equilibrium.

Next, we can consider a slight generalization of the foregoing discussion, which leads to the Boltzmann distribution for systems having such a widely spaced set of allowed levels that they cannot be handled by means of a continuous $C(w)$ function.

We consider a composite system as before, which consists of N molecules in a box and in addition, the arbitrary quantized system whose allowed energy levels will be written $w_1, w_2, \cdots, w_a, \cdots$ with the corresponding statistical weights (order of degeneracy) $g_1, g_2, \cdots, g_a$. We suppose that the total energy of the composite system is known to lie between W and $W+\delta W$ and ask for the probability that the quantized part be found in the αth energy level, assuming that all states of the composite system with total energy between W and $W+\delta W$ are equally probable.

The probability of finding the quantized part in the αth energy level will, therefore, be proportional to

$$P(W_a)\sim C_N'(W-w_a)g_a\delta W,$$

which is also

$$P(w_a)\sim(1-w_a/W)^{3N/2-1}g_a\delta W.$$

As before, we may introduce the energy kT defined by (7) and recognize that the first factor is essentially equal to $e^{-w_a/kT}$ if N is a very large number. Therefore, we have

$$P(w_a)\sim g_ae^{-w_a/kT}, \quad (11)$$

which is the familiar Boltzmann distribution law for systems having quantized energy levels. As before, we recognize the fact that the probability is proportional to g_a, the number of states of the quantized system of energy w_a, and the exponential factor representing the dependence on w_a of the density of states of the N molecules in a box which constitute the perfect gas thermometer-thermostat with which the quantized system is in equilibrium.

With the derivation of the Boltzmann law accomplished, one can proceed to develop the theory of the thermal properties of matter in the usual way by introducing the partition function

$$z(T)=\sum_{\alpha} g_{\alpha} e^{-w_{\alpha}/kT}, \tag{12}$$

from which the mean energy at temperature T is calculated by the formula

$$\bar{w}(T)=-k\frac{d\log z}{d(1/T)}. \tag{13}$$

The uses of the partition function are so well known that it will not be necessary here to repeat that part of the development.

In deriving the distribution law for the distribution of translational energy of a single molecule in equilibrium with N molecules, the assumption was made that the energy of the $(N+1)$ molecules is precisely known to be W. Looking back over the argument leading to (8), we see that the argument would have been essentially unaltered had we assumed that the composite system's energy had a value between W and $W+\delta W$.

Actually, when we have a gas at temperature T it will have a distribution-in-energy of its total energy, which we may derive by assuming the N molecules in a box to be part of a much larger composite system which includes N' more molecules where N' is very large compared to N. In that case we find that the distribution in energy of the N molecules is given by

$$P(w)dw\sim e^{-w/kT}w^{3N/2-1}dw, \tag{14}$$

where kT is defined by $W=3N'kT/2$ which is supposed to be negligibly different from $3(N'+N)kT/2$ since $N'\gg N$. From (14) we may calculate the mean energy of the system of N molecules in thermal equilibrium with the larger system. It comes out

$$\bar{w}=3NkT/2, \tag{15}$$

which justifies the identification of T with the usual absolute temperature on the perfect gas scale. Similarly one may calculate the mean-square deviation from the mean of the energy of the N molecules

$$\Delta^2=(w-w_{Av})^2{}_{Av}=3N(kT)^2/2. \tag{16}$$

Hence the fractional fluctuation in energy of the N molecules is

$$\Delta/w_{Av}=(3N/2)^{-\frac{1}{2}}, \tag{17}$$

which is extremely small when N is large. This justifies the neglect of the distribution in energy of the N molecules in the argument leading to (8) by which the distribution in energy of one molecule was found.

The distribution law for the case of Fermi-Dirac statistics[2] is easily obtained as follows. We consider the system of N equivalent particles to which the Fermi statistics is to be applied, as a single quantized system in equilibrium with a larger perfect gas-thermometer thermostat. Then the distribution in energy of its quantized states is given by application of (11). In other words the relative probability of each of the independent states of the N equivalent particles governed by the Pauli exclusion principle is given by the Boltzmann factor, $e^{-w/kT}$.

The exclusion principle tells us that no two particles can be in the same quantum state and, therefore, if the allowed energy levels of a single particle are given by w_{α}, each allowed level of the system of N particles will be characterized by a set of quantum numbers N_{α}, one for each single particle state α, where N_{α} can have only the values 0 and 1 and $\Sigma_{\alpha}N_{\alpha}=N$. The total energy is

$$W=\sum_{\alpha} w_{\alpha}N_{\alpha}.$$

To find the distribution-in-energy of the single particle in the case of Fermi-Dirac statistics, we have to calculate the mean value of N_{β}, the probability of occupation of the βth state by a particle. By the Boltzmann principle this is

$$(N_{\beta})_{Av}=\frac{\sum_0 0e^{-W/kT}+\sum_1 1e^{-W/kT}}{\sum_{0+1}e^{-W/kT}},$$

where Σ_0 means the sum over all sets of N's consistent with $\Sigma N_{\alpha}=N$ and having $N_{\beta}=0$ while Σ_1 means summation over all sets of N's having $\Sigma N_{\alpha}=N$ and having $N_{\beta}=1$. If we write

$$W'=W-N_{\beta}w_{\beta},$$

this can be written

$$(N_{\beta})_{Av}=\frac{e^{-w_{\beta}/kT}\sum_1 e^{-W'/kT}}{\sum_0 e^{-W'/kT}+e^{-w_{\beta}/kT}\sum_1 e^{-W'/kT}}$$

[2] A valuable account of Fermi-Dirac gas theory, together with its most important field of application, the Sommerfeld electron theory of metals, is given in the article by Sommerfeld and Bethe, *Handbuch der Physik*, Vol. 24/2 (Julius Springer, Berlin, 1933), p. 333.

and, therefore,

$$(N_\beta)_N = \frac{1}{A(T)e^{w_\beta/kT}+1}, \tag{18}$$

in which $A(T) = \sum_0 e^{-w'/kT} / \sum_1 e^{-w'/kT}$. (19)

Here the parameter $A(T)$ is the ratio of the partition function for an N particle system from which the state w_β is excluded from the set of allowed single particle states to that for an $(N-1)$ particle system from which this same state is excluded.

Evidently, if $A(T) \gg 1$, then the $+1$ in the denominator of (18) will be negligible and $(N_\beta)_N$ will be approximately equal to $e^{-w_\beta/kT}$ times a factor that is independent of w_β. Therefore, in this limit the distribution law is not appreciably affected by the operation of the Pauli exclusion principle. On the other hand, if $A(T) \ll 1$ then for values of w_β small enough that $A(T)e^{w'_\beta/kT} \ll 1$, we shall have $(N_\beta)_N = 1$, that is, the low energy states are almost certainly occupied by one particle in each state which means a large departure from the classical distribution law.

The probability that a single particle have energy between w and $w+dw$ is, therefore, proportional to the product of the number of states in this energy range, $c'(w)dw$ multiplied by the chance that a single particle state of energy w be occupied which is $[Ae^{w/kT}+1]^{-1}$ so the distribution function is

$$P(w)dw \sim c'(w)[Ae^{w/kT}+1]^{-1}dw, \tag{20}$$

which is the familiar form from which the usual deductions of properties of the Fermi-Dirac gas may be made.

In conclusion, it is fitting to remark on the occasion of the seventieth birthday of one of the greatest teachers and productive workers that modern theoretical physics has known—Arnold Sommerfeld. Everyone of my generation grew up on atomic physics by way of his great *Atombau und Spektrallinien*, a large group have profited by the stimulation of his lectures on his American visits, and a fortunate few of us have derived boundless stimulation from the opportunity of working in his Institut für theoretische Physik in the former brighter days. All physicists join in wishing him a happy birthday and continued vigor with which to participate in the further developments of fundamental ideas to which he has contributed so extensively.

DECEMBER 1, 1938 PHYSICAL REVIEW VOLUME 54

Transitions Between Levels Spaced Almost Continuously

ALFRED LANDÉ
Mendenhall Laboratory, Ohio State University, Columbus, Ohio
(Received June 16, 1938)

The transition probabilities between closely spaced energy levels, when calculated according to the usual perturbation methods, lead to apparent infinities if the perturbation is finite. A more appropriate method of approximation which disposes of the infinities is worked out. The resulting finite transition probabilities are identical with the finite parts of the usual expressions in first and second order, but additional finite terms appear in orders higher than the second.

THE probabilities of transition between the levels of a continuous or almost continuous energy spectrum can always be treated in an unambiguous way as far as the first-order (direct) transitions are concerned. But the second-order transitions (through *one* intermediate state) give rise to infinite transition probabilities in the case of a continuous spectrum. Usually one splits up the resulting terms, in a more or less arbitrary way, into a finite part that is supposed to represent the physical facts, and an infinite part that is neglected without further justification. This procedure appears all the more ambiguous when one learns that different approximation methods lead to different convergent parts of diverging series. The omission of infinite terms seems to be based on the hope that the various infinite members would cancel one another if

only a proper way of approximation were introduced.[1] In the following we propose a method of approximation that is more appropriate to the case of a continuous energy spectrum. Furthermore, whereas the perturbation method is confined to perturbations that are small compared with the unperturbed intervals, we here strive to do away with this nonphysical restriction in the continuous case. The results obtained in this way turn out to be free from the infinities mentioned before. They allow one at the same time to trace down the origin of the infinities that appear in the usual perturbation method.

§1.

In order to show the divergence in question we first use an expansion with respect to the perturbation time τ. If

$$\Psi^0(x, t)=\sum_l a_l\psi_l{}^0(x)\exp[iE_l{}^0t/\hbar]$$

is an unperturbed ψ-function, and if a perturbing potential V is applied there will be a change of the coefficients a_l with the time

$$\dot{a}_l=\sum_k a_k(t)iv_{kl}\exp[i\omega_{kl}t], \tag{1}$$

where $\omega_{kl}=(E_k{}^0-E_l{}^0)/\hbar$

and

$$v_{kl}=(1/\hbar)\int\psi_k{}^0V\psi_l{}^{*0}dx. \tag{2}$$

One can try to find the a_l after the time τ by means of a Taylor series

$$a_l(\tau)=a_l(0)+\tau\dot{a}_l(0)+(\tau^2/2!)\ddot{a}_l(0)+\cdots.$$

If one inserts on the right for $\dot{a}_l(0)$, $\ddot{a}_l(0)$ the values obtained from (1) and its derivatives, and assumes $a_n(0)$ to be unity and all the other $a_l(0)$ to be zero, one obtains the following power series with respect to the time τ:

$$a_l(\tau)=\frac{i\tau}{1!}v_{nl}+\frac{(i\tau)^2}{2!}[v_{nl}\omega_{nl}+\sum_j v_{nj}v_{jl}]+\frac{(i\tau)^3}{3!}[v_{nl}\omega_{nl}{}^2+\sum_j v_{nj}v_{jl}\omega_{nl}+\sum_i\sum_j v_{nj}v_{ji}v_{il}]+\cdots. \tag{3}$$

[1] See L. Nordheim's remarks in *Théorie des chocs et du rayonnement* (Annales Institut Poincaré, 1936), p. 8.

This series can be rearranged in various ways. For instance, when transitions from $E_n{}^0$ to *neighboring levels* $E_l{}^0$ with small ω_{nl} are considered, one may rearrange (3) into a series with respect to powers of ω_{nl}. The resulting formula is not written down explicitly. When the *perturbation* is small one may rearrange (3) with respect to powers of V:

$$\left(\frac{\exp[i\omega_{nl}\tau]-1}{\omega_{nl}}\right)v_{nl}+\left(\frac{\exp[i\omega_{nl}\tau]-1}{\omega_{nl}{}^2}-\frac{i\tau}{\omega_{nl}}\right)\times\sum_j v_{nj}v_{jl}+\cdots. \tag{4}$$

The three power series with respect to τ, ω_{nl} and V do not represent the same values, however, because $a_l(\tau)$ does not converge *uniformly* in τ, ω_{nl} and V. In the most important applications τ as well as ω_{nl} and V are small, and the result will then entirely depend on the limiting process applied. In the series (4) the first power of V is multiplied by an exponential factor that contains already the highest powers of τ and ω_{nl}. In the case of radiation one is interested in finding the total probability $J(\tau)$ of transitions from an initial level n into a *range* of levels l distributed with the spectral density $\rho(E_l{}^0)$:

$$J(\tau)=\int|a_l(\tau)|^2\cdot\rho(E_l{}^0)\cdot dE_l{}^0. \tag{5}$$

When using the first term (4) alone one obtains the well-known result

$$J(\tau)=\int\frac{\sin^2(\omega_{nl}\tau)}{\omega_{nl}{}^2}|v_{nl}|^2\rho(E_l{}^0)dE_l{}^0=h\pi\tau\cdot\rho(E_n{}^0)\cdot|v_{nl}|^2{}_{l\to n} \tag{5'}$$

where the matrix element v_{nl} refers to a transition from $E_n{}^0$ to a level $E_l{}^0$ of the *same* energy. This first approximation becomes questionable, however, when one sees that the higher terms of (4) contain denominators $\omega_{nl}{}^2$ which lead, when integrated in (5) over $dE_l{}^0$, to *infinities*. On the other hand $\sum_j$ in the second term of (4) will tend toward *zero* if the perturbation becomes small. In the limit when both ω_{nl} and V become small, one arrives at an indefinite product $\infty\cdot 0$, which contributes to the integral (5). In the case of a finite perturbation, but ever decreasing

unperturbed intervals, the value of this product would approach ∞. This is just one example of the insufficiency of the perturbation method. Other series than the above for $a_l(\tau)$ with different divergent results (see below §4) have been used by various authors. Our aim is to get rid of the infinities altogether.

§2.

In order to obtain a convergent expression for the transition probability in the case of an almost continuous spectrum, we first derive a rigorous expression for $a_l(\tau)$ for any value of τ and for any magnitude of V. Approximations will be left to a later stage of the calculations (§3).

Up to the time 0 and after the time τ we have the unperturbed potential U^0 with eigenfunctions $\psi_k{}^0$. From 0 to τ the potential U^0+V gives rise to eigenfunctions $\psi_{m'}$ and eigenvalues $E_{m'}$. Before $t=0$ only the level $E_n{}^0$ shall be present with amplitude 1. We can then expand, at $t=0$, the unperturbed ψ-function into a series of perturbed ψ-functions:

$$\psi_{(x,\,0)}=\psi_n{}^0(x)=\sum_m c^*_{m'n}\psi_{m'}(x), \tag{6}$$

whose coefficients $c_{m'n}{}^*$ are defined by

$$\psi_{m'}=\sum_l c_{m'l}\psi_l{}^0 \quad \text{or} \quad c_{m'l}=\int \psi_{m'}\psi_l{}^{0*}dx. \tag{7}$$

By the time τ the wave function has become

$$\psi_{(x,\,\tau)}=\sum_{m'} c^*_{m'n}\psi_{m'}(x)\cdot\exp\,[iE_{m'}\tau/\hbar]. \tag{7'}$$

Expanding this last function into the series

$$\psi_{(x,\,\tau)}=\sum a_l(\tau)\cdot\psi_l{}^0(x)\cdot\exp\,[iE_l\tau/\hbar]$$

we find after equating the right-hand sides of the last two equations:

$$a_l(\tau)=\int \psi_{(x,\,\tau)}\cdot\psi^{0*}(x)\exp\,[-iE_l{}^0\tau/\hbar]dx.$$

Using (6) and (7) we obtain

$$a_l(\tau)=\sum_{m'}\exp\,[i(E_{m'}-E_l{}^0)\tau/\hbar]c_{m'l}c^*_{m'n}$$

for every l including $l=n$ and for every V. For $\tau=0$ we have in particular

$$a_l(0)=\sum_{m'}c_{m'l}c^*_{m'n} \text{ which is } \delta_{nl}=\begin{matrix}1 & l=n\\ 0 & l\neq n\end{matrix} \text{ for } \tag{7''}$$

according to the completeness relation. Thus $a_l(\tau)$ for $l\neq n$ can also be written

$$a_l(\tau)=\sum_{m'}(\exp\,[i\langle E_{m'}-E_l{}^0\rangle\tau/\hbar]-1)\cdot c_{m'l}c^*_{m'n}. \tag{8}$$

The $c_{m'l}$ may be expressed in terms of the perturbation. Multiplying the wave equation for $\psi_{m'}$ by $\psi_l{}^{0*}$ and the wave equation for ψ_l by $\psi_{m'}$, subtracting the products and integrating over the space coordinates we obtain

$$c_{m'l}(E_{m'}-E_l{}^0)=V_{m'l} \tag{9}$$

where $$V_{m'l}=\int\psi_{m'}V\psi_l{}^{0*}dx.$$

Thus, if $E_{m'}\neq E_l{}^0$, one has $c_{m'l}=V_{m'l}/(E_{m'}-E_l{}^0)$. Next we show that $c_{m'l}$ is proportional to $c_{m'n}$. From (9) we have first

$$c_{m'l}=\frac{V_{m'l}}{E_{m'}-E_l{}^0} \text{ with } V_{m'l}=\int\psi_{m'}V\psi_l{}^{0*}dx. \tag{10}$$

Under the integral we can replace $\psi_{m'}$ by the series

$$\psi_{m'}=c_{m'n}\psi_n{}^0+\sum_j{}'c_{m'j}\psi_j{}^0 \tag{10'}$$

where the sum is extended only over $j\neq n$. Each $c_{m'j}$ under the sum $\sum'$ can be replaced by a quotient as in (10), and the same process can be repeated. This iteration leads to the proportionality

$$c_{m'l}=\frac{c_{m'n}}{E_{m'}-E_l{}^0}\cdot V_{nl}{}^{(m')}, \tag{11}$$

where

$$V_{nl}{}^{(m')}=V_{nl}+\sum{}'_{j\neq n}\frac{V_{nj}V_{jl}}{E_{m'}-E_j{}^0}$$

$$+\sum_{i\neq n}{}'\sum_{j\neq n}{}'\frac{V_{ni}V_{ij}V_{jl}}{(E_{m'}-E_i{}^0)(E_{m'}-E_j{}^0)}+\cdots \tag{12}$$

and where the V_{ij} are defined by

$$V_{ij}=\int\psi_i{}^0V\psi_j{}^{0*}dx.$$

Inserting this value of $c_{m'l}$ into (8) we obtain

$$a_l(\tau)=\sum_{m'}(\exp\,[i(E_{m'}-E_l{}^0)\tau/\hbar]-1)\cdot\frac{V_{nl}{}^{(m')}}{E_{m'}-E_l{}^0}\cdot|c_{m'n}|^2. \tag{13}$$

All formulae so far are rigorous.

§3.

We now are going to carry out some approximations. Because of the factor (10)

$$|c_{m'n}|^2=\left|\frac{V_{m'n}}{E_{m'}-E_n^0}\right|^2,$$

the members m' of the sum (13) have a maximum for those m' whose energy $E_{m'}$ is close to E_n^0. If the perturbation happens to be of such a magnitude that *one* perturbed level $E_{m'}$ is found to be very close to E_n^0 (much closer than to $E^0{}_{n+1}$ or $E^0{}_{n-1}$) then there is a very steep maximum for this particular m' which shall be called N' hereafter. If the perturbation V is varied slightly so as to shift $E_{N'}$ to a position say halfway between E_n^0 and $E^0{}_{n+1}$, then the maximum will be only moderately steep, and a number of members m' besides N' will contribute to the sum (13). We now introduce the *assumption* that in general, barring singular cases of discontinuous potential functions, the value of the sum (13) will not depend on such small variations of the magnitude of V that would shift $E_{N'}$ over an interval between E_n^0 and $E^0{}_{n+1}$. That is to say, we assume that the *value* of (13) does not depend on whether the magnitude of V is varied slightly such as to produce a very steep instead of a moderately steep maximum under the sum. We are then allowed to *calculate* the value of (13), in the case of an almost continuous energy spectrum, for such a magnitude of V that *one* $E_{m'}$ which is called $E_{N'}$ is very close to E_n^0 and thus contributes alone to the sum (13). This reduces (13) to

$$a_l(\tau)=\frac{\exp\left[i(E_{N'}-E_l^0)\tau/\hbar\right]-1}{E_{N'}-E_l^0}\cdot V_{nl}{}^{(N')} \quad (14)$$

where $E_{N'}\sim E_n^0$ and where

$$V_{nl}{}^{(N')}=V_{nl}+\sum{}'_{j\neq n}\frac{V_{nj}V_{jl}}{E_{N'}-E_j^0}+\sum{}_i'\sum{}_j'\frac{V_{ni}V_{ij}V_{jl}}{(E_{N'}-E_i^0)(E_{N'}-E_j^0)}+\cdots. \quad (14')$$

The value of (14), according to our assumption, is independent of whether the perturbed level $E_{N'}$ is *very* close to or actually coincides with E_n^0. Thus we can finally write, without physical restriction, for $l\neq n$

$$a_l(\tau)=\frac{\exp\left[i(E_n^0-E_l^0)\tau/\hbar\right]-1}{E_n^0-E_l^0}\cdot V_{nl}{}^{(n)}, \quad (15)$$

where

$$V_{nl}{}^{(n)}=V_{nl}+\sum{}'_{j\neq n}\frac{V_{nj}V_{jl}}{E_n^0-E_j^0}+\sum{}_i'\sum{}_j'\frac{V_{ni}V_{ij}V_{jl}}{(E_n^0-E_i^0)(E_n^0-E_j^0)}+\cdots. \quad (15')$$

The first-order term of (15′) is identical with that of (14′); the higher terms of (15′) and (14′) agree if our assumption is correct. Inserting (15) in the integral (5) one obtains

$$J(\tau)=\pi\tau\hbar\cdot\rho(E_n^0)\cdot|(1/\hbar)V_{nl}{}^{(n)}|^2{}_{E_l^0\to E_n^0}. \quad (16)$$

In contrast to (4), our result (15) (15′) does not contain denominators $E_l^0-E_n^0$ in higher than first order, *the integral over $a_l(\tau)$ will not give rise to infinities.*

§4.

Comparison with other solutions. The expression (4) for $a_l(\tau)$ contains denominators $\omega_{nl}=E_n^0-E_l^0$ and also higher powers of ω_{nl} which give rise to infinities in the integral $J(\tau)$. The same infinities reappear in the well-known result[2] given by the perturbation theory:

$$V_{nl}+\left(\sum{}'_{j\neq n}\frac{V_{nj}V_{jl}}{E_n^0-E_j^0}-\frac{V_{nn}V_{nl}}{E_n^0-E_l^0}\right)+\cdots. \quad (17)$$

The second-order term is identical, in its first part, with our expression (15′). Its second part, however, contains the fatal denominator ω_{nl} which leads to an infinity in $J(\tau)$. The origin of this term as the result of the perturbation method can be traced down as follows. Going back to the equations (14) and (14′) and applying them to so small a perturbation V that $E_{N'}$ is the perturbed level of E_n^0 itself (instead of that of a distant unperturbed level) one may expand the denominators $(E_{N'}-E_l^0)$ in terms of

[2] See E. C. Kemble, *Quantum Mechanics* (McGraw-Hill, 1937), p. 386, formula (47.27).

the unperturbed levels, that is, one may write in (14)

$$V_{nl}^{(N')}\frac{1}{E_{N'}-E_l^0}=\left(V_{nl}+\sum{}'_{j\neq n}\frac{V_{nj}V_{jl}}{E_{N'}-E_l^0}+\cdots\right)\cdot\left(\frac{1}{E_n^0-E_l^0}-\frac{E_{N'}-E_n^0}{(E_n^0-E_l^0)^2}+\cdots\right).$$

If we neglect all terms of higher than second order, and notice that $E_{N'}-E_n^0=V_{nn}$ because of (9) and (7), we obtain

$$\frac{1}{E_n^0-E_l^0}\left[V_{nl}+\sum{}'_j\frac{V_{nj}V_{jl}}{E_n^0-E_l^0}-\frac{V_{nn}V_{nl}}{(E_n^0-E_l^0)}+\cdots\right];$$

hence according to (14)

$$a_l(\tau)=\frac{\exp\left[i(E_{N'}-E_l^0)\tau/h\right]-1}{E_n^0-E_l^0}\cdot\left[V_{nl}+\left(\sum{}'_j\frac{V_{nj}V_{jl}}{E_n^0-E_l^0}-\frac{V_{nn}V_{nl}}{E_n^0-E_l^0}\right)+\cdots\right].$$

The square bracket turns out to be identical with (17), whereas the time factor before the bracket contains $E_{N'}$ in its exponent and E_n^0 in its denominator, instead of $E_{N'}$ in both places as in (14) or E_n^0 in both places as in (15). The nonconvergent term in the usual expression (17) is thus found to originate from an inconsistency of approximation: If in (14), when passing over to (15) one would replace $E_{N'}$ by E_n^0 only in the denominator but not in the exponent, then $a_l(\tau)$ would converge toward const/0$=\infty$ instead of toward 0/0= finite for $E_l^0\to E_n^0$.

The divergent series (4) is correct only in its first-order convergent term. The divergent series (17) agrees with our result only in its first- and second-order terms and only when the divergent part of the latter is dropped. Our third- and higher-order terms differ however from the usually given expressions[1] which read in third order

$$\sum{}'_i\sum{}'_j\frac{v_{ni}v_{ij}v_{jl}}{\omega_{ni}\omega_{ij}}$$

in contrast to our

$$\sum{}'_i\sum{}'_j\frac{v_{ni}v_{ij}v_{jl}}{\omega_{ni}\omega_{nj}}.$$

The latter can be transformed by virtue of

$$\frac{1}{\omega_{ni}\omega_{nj}}=\frac{1}{\omega_{ni}\omega_{ij}}+\frac{1}{\omega_{nj}\omega_{ji}}$$

for $E_n^0\to E_l^0$ into the more symmetrical form

$$\sum{}'_i\sum{}'_j\left(\frac{v_{ni}v_{ij}v_{jl}}{\omega_{ni}\omega_{ij}}+\frac{v_{ni}v_{ij}v_{jl}}{\omega_{ij}\omega_{jl}}\right),\qquad(18)$$

whose first part is identical with the usually given third-order term. Not so, however, the second part of (18) since the state n is different from the state l although $E_n^0=E_l^0$. The customary expression[3] is asymmetric in so far as it contains in its denominator the interval between the initial and the first intermediate level but lacks the interval between the last intermediate and the final level.

A few remarks may be added about *degeneracy*. In general V_{nl} does not vanish. If, however, for a group of final states l in the neighborhood of n, all the V_{nl} happen to vanish like $(E_n^0-E_l^0)$, then we have degeneracy: The group l together with n may be considered as having had originally exactly the same energies, then as being separated first by a potential, say $(1/10)V$, to which is added later the perturbation $(9/10)V$. The relation $V_{nl}=0$ means then that V is *adapted* to the originally coinciding states in so far as both the integrals

$$\int\psi_n{}^0\psi_l{}^{0*}dx\quad\text{and}\quad\int V\cdot\psi_n{}^0\psi_l{}^{0*}dx$$

vanish. If V_{nl} vanishes like $(E_n^0-E_l^0)^2$ then the second-order term in (15′) will also vanish, and the series begins with the third-order term. The difference between our formula and the customary expression would then have physical significance.

[3] See for instance reference 1, p. 10, formula (3.8)

A Solution of the Schrödinger Equation by a Perturbation of the Boundary Conditions

H. FROEHLICH
H. H. Wills Physical Laboratory, University of Bristol, England
(Received September 29, 1938)

A perturbation method is developed which allows a simple calculation of the eigenvalues of the Schrödinger equation by a slight change in the boundary condition.

§1

IN recent years the solution of the Schrödinger equation with a boundary condition at the surface of a finite space has become of some importance. In Wigner and Seitz' theory of metallic wave functions,[1] for instance, $\partial\psi/\partial r$ must vanish at the surface of a sphere. An example with the boundary condition $\psi=0$ at the surface of a sphere was given by Sommerfeld and Welker.[2] In most of these problems we want to calculate the eigenvalue E as a function of certain parameters which determine the boundary conditions. In the above examples, for instance, we want E as a function of the radius of the sphere. Now let us assume that we have solved the Schrödinger equation for a given radius r_0. Then, one should imagine that a calculation of the eigenvalue for a somewhat different radius r_0' is possible by means of a perturbation method without giving an exact solution of the Schrödinger equation for this new boundary condition. The ordinary Schrödinger perturbation method cannot be used in the present case, since we do not deal with a perturbation of the potential, but with a perturbation of the boundary condition. We therefore have to find a different method. The special case of metallic wave functions has been already treated in a previous paper.[3] In the following we want to describe a more general method[4] for the solution of the Schrödinger equation by a perturbation of the boundary condition.

§2

Let E_0 be a nondegenerate eigenvalue and ψ_0 the wave function, satisfying the Schrödinger equation (we use atomic units)

$$\Delta\psi_0+(E_0-V)\psi_0=0 \tag{1}$$

with a certain boundary condition at the surface S_0. The wave function ψ_0 can be chosen as a real function since E_0 is not degenerate. We want to calculate the eigenvalue $E=E_0+\epsilon$ satisfying the Schrödinger equation

$$\Delta\psi+(E_0+\epsilon-V)\psi=0 \tag{2}$$

with a slightly different boundary condition at a slightly different surface[5] S.

We multiply Eq. (1) with ψ and (2) with ψ_0 and subtract both equations. We then find after integration over the space bounded by S with the use of Green's theorem:

$$\int_S\left(\psi_0\frac{\partial\psi}{\partial n}-\psi\frac{\partial\psi_0}{\partial n}\right)d\sigma+\epsilon\int\psi_0\psi d\tau=0. \tag{3}$$

$d\sigma$ means integration over the surface S; n is the normal to S in the outward direction. We shall consider the following four boundary conditions:

A: $\psi_0=0$ at S_0, $\psi=0$ at S (4a)

B: $\psi_0=0$ at S_0, $\psi=\mu F$ at S, $S=S_0$ (4b)

C: $\partial\psi_0/\partial n=0$ at S_0, $\partial\psi/\partial n=0$ at S, (4c)

D: $\partial\psi_0/\partial n=0$ at S_0, $\partial\psi/\partial n=\mu G$ at S, $S=S_0$, (4d)

where μ is small parameter. The functions F and G are known at the surface S_0. According to (3) we find for:

A: $$N\epsilon=-\int_S\psi_0\frac{\partial\psi}{\partial n}d\sigma, \tag{5a}$$

B: $$N\epsilon=\mu\int_{S=S_0}F\frac{\partial\psi_0}{\partial n}d\sigma, \tag{5b}$$

[1] Wigner and Seitz, Phys. Rev. **46**, 509 (1934).
[2] Sommerfeld and Welker, Ann. d. Physik **32**, 56 (1938).
[3] H. Froehlich, Proc. Roy. Soc. **58**, 97 (1936); and *Elektronentheorie der Metalle* (Berlin 1936), p. 272. Cf. also the discussion by Bardeen, J. Chem. Phys. **6**, 372 (1938).
[4] A different method was proposed by L. Brillouin, Comptes rendus **204**, 1863 (1938).
[5] This includes $S=S_0$, or the case of the same boundary condition at S and S_0 but $S\neq S_0$.

C: $$N\epsilon=\int_S \psi\frac{\partial\psi_0}{\partial n}d\sigma, \tag{5c}$$

D: $$N\epsilon=-\mu\int_{S=S_0}\psi_0 G d\sigma, \tag{5d}$$

with $$N=\int_S \psi_0\psi d\tau.$$

§3

First approximation

The above expressions (5a)–(5d) have been obtained without any approximations. We shall now make use of the fact that the change in the boundary condition is small, i.e., that the functions ψ and ψ_0 must be nearly equal. We therefore can put for N

$$N=\int_S \psi_0\psi d\tau \backsimeq \int_S \psi_0^2 d\tau \backsimeq \int_{S_0}\psi_0^2 d\tau = 1.$$

Furthermore, in Eq. (5a) we may replace $\partial\psi/\partial n$ by $\partial\psi_0/\partial n$ and in Eq. (5c) ψ by ψ_0.[4, 6] We are now able to calculate ϵ in terms of the known functions ψ_0, F or G:

A: $$\epsilon=-\int_S \psi_0\frac{\partial\psi_0}{\partial n}d\sigma, \tag{6a}$$

B: $$\epsilon=\mu\int_{S=S_0} F\frac{\partial\psi_0}{\partial n}d\sigma, \tag{6b}$$

C: $$\epsilon=\int_S \psi_0\frac{\partial\psi_0}{\partial n}d\sigma, \tag{6c}$$

D: $$\epsilon=-\mu\int_{S=S_0} G\psi_0 d\sigma. \tag{6d}$$

Thus, we see that it is possible to determine the first approximation of the energy $E=E_0+\epsilon$ by an integration over functions which are known from the zero approximation (ψ_0, $\partial\psi_0/\partial n$) or from the boundary conditions (μF, μG).

[6] It should be noticed that an analogous proceeding could not be allowed if we had chosen S_0 instead of S as the boundary for the integration in (3). In that case the right-hand side in Eq. (5a), for instance, would be replaced by

$$\int_{S_0}\psi\frac{\partial\psi_0}{\partial n}.$$

Here we cannot insert ψ_0 instead of ψ since $\psi_0=0$ at S_0.

§4

Examples

The expressions for ϵ in the cases A and C may still be simplified considerably if $\psi_0=\psi_0(r)$ is spherically symmetrical and S_0 is the surface of a sphere with radius r_0.

Let $\mathbf{r}_1$ be the radius vector from the origin of the sphere to a point at the surface S_1. Since S is assumed to be only slightly different from S_0, we have

$$||\mathbf{r}_1|-r_0|\ll r_0 \tag{7}$$

and $$|\sin\alpha|\ll 1, \tag{8}$$

where α is the angle between $\mathbf{r}_1$ and the normal $\mathbf{n}$ to S at the point where $\mathbf{r}_1$ reaches S. We therefore may develop ψ_0 and $\partial\psi_0/\partial n$, using the fact that ψ_0 is spherically symmetrical:

$$(\psi_0(\mathbf{r}_1))_S=(\psi_0)_{S_0}+(\partial\psi_0/\partial r)_{S_0}\times(|\mathbf{r}_1|-r_0)+\cdots, \tag{9}$$

$$\left(\frac{\partial\psi_0(\mathbf{r}_1)}{\partial n}\right)_S=\left(\frac{\partial\psi_0(\mathbf{r}_1)}{\partial r}\right)_S\frac{\partial r_1}{\partial n}=\left\{\left(\frac{\partial\psi_0}{\partial r}\right)_{S_0}+\left(\frac{\partial^2\psi_0}{\partial r^2}\right)_{S_0}(|\mathbf{r}_1|-r_0)\right\}\cos\alpha+\cdots. \tag{10}$$

Case A:—From (6a), we find in the first approximation (i.e., neglecting terms with higher power in $(|\mathbf{r}_1|-r_0)$) using (9), (10), (4a) and (7):

$$\epsilon=-\left(\frac{\partial\psi_0}{\partial r}\right)^2_{S_0}\int(|\mathbf{r}_1|-r_0)\cos\alpha d\sigma = -\left(\frac{\partial\psi_0}{\partial r}\right)^2_{S_0}v. \tag{11}$$

Here, regarding Eq. (8), v is the difference of the volumes bounded by S and S_0 respectively. v is positive if S includes a greater volume than S_0.

Case C:—In this case we find from (6c) in the first approximation, using (9), (10), (4c) and (7):

$$\epsilon=(\psi_0\partial^2\psi_0/\partial r^2)_{S_0}v,$$

where v is the same volume as defined above. Using the Schrödinger equation (1) and Eq. (4c), we find

$$(\partial^2\psi_0/\partial r^2)_{S_0}=[V(r_0)-E_0]\psi_0(r_0)$$

and, therefore, since $(\psi_0)s_0=\psi_0(r_0)$

$$\epsilon=[V(r_0)-E_0]\psi_0^2(r_0)v. \quad (12)$$

From Eqs. (11) and (12) we see that the first-order perturbation energy is zero if the sphere S_0 is only deformed without a change of volume.

§5

Higher approximations

The higher approximations cannot usually be calculated without solving a partial differential equation.

We do not intend to go into any detail of this question. We just want to mention that it is possible to eliminate the potential from the Schrödinger equation. In order to demonstrate that, we proceed thus: We insert

$$\psi=\psi_0 f$$

into Eq. (2) and observe that ψ_0 obeys Eq. (1). We then obtain for f the differential equation:

$$\text{div}\,(\psi_0^2\,\text{grad}\,f)+\epsilon\psi_0^2 f=0.$$

In cases where the function ψ_0 has a simple form, this transformation may be of advantage.[3]

DECEMBER 1, 1938 PHYSICAL REVIEW VOLUME 54

On the Bose-Einstein Condensation

F. LONDON
*Institut Henri Poincaré, University of Paris, France**

(Received October 12, 1938)

A proof is given of the condensation phenomenon of a Bose-Einstein gas. A preliminary discussion of its transport properties is outlined with a view to its possible bearing on the problem of liquid helium.

INTRODUCTION

IN his well-known papers[1] on the degeneracy of an ideal gas, Einstein mentioned a peculiar condensation phenomenon of the ideal "Bose-Einstein" gas. This very interesting discovery, however, has not appeared in the textbooks, probably because Uhlenbeck in his thesis[2] questioned the correctness of Einstein's argument. Since, from the very first, the mechanism appeared to be devoid of any practical significance, all real gases being condensed at the temperature in question, the matter has never been examined in detail; and it has been generally supposed that there is no such condensation phenomenon.

In discussing some properties of liquid helium, I recently realized that Einstein's statement has been erroneously discredited; moreover, some support could be given to the idea that the peculiar phase transition ("λ-point"), that liquid helium undergoes at 2.19°K, very probably has to be regarded as the condensation phenomenon of the Bose-Einstein statistics, distorted, of course, by the presence of molecular forces and by the fact that it manifests itself in the liquid and not in the gaseous state. In a preliminary note,[3] the course of the specific heat of an ideal Bose-Einstein gas was reproduced, but no proof was communicated. As, since then, I have been asked several times for a proof and as even the correctness of the result has been questioned anew, it might perhaps be justified, on this occasion to publish a quite elementary demonstration of the condensation mechanism, discussing only briefly here the possible connection of the Bose-Einstein degeneracy with the problem of liquid helium.[4]

* At present at Duke University, Durham, North Carolina.

[1] A. Einstein, Ber. Berl. Akad. 261 (1924); 3 (1925).

[2] G. E. Uhlenbeck, *Dissertation* (Leiden, 1927).

[3] F. London, Nature **141**, 643 (1938).

[4] In a recent paper Uhlenbeck has withdrawn his former objection. G. E. Uhlenbeck and B. Kahn, Physica **5**, 399 (1938).

§1. The Degeneracy of the Bose-Einstein Gas

We start from the well-known fundamental formula for the most probable distribution in a Bose-Einstein gas:

$$N_i=\frac{g_i}{e^{\beta\epsilon_i+\alpha}-1}.$$

Here $\beta=1/kT$, g_i=statistical weight of the state of energy ϵ_i; the parameter α has to be determined as a function of T by the condition

$$N=\sum_i N_i=\sum_i \frac{g_i}{e^{\beta\epsilon_i+\alpha}-1}. \qquad (1)$$

N=total number of particles. It may be mentioned that α is proportional to Gibb's potential ζ; it is $\zeta=-\alpha kT$. For free particles of the mass M, without spin, in a given volume V, one usually assumes

$$\epsilon_p=(p^2/2M),\quad g(p)dp=(4\pi V/h^3)p^2dp$$

and obtains

$$N=\frac{4\pi V}{h^3}\int_0^\infty \frac{p^2dp}{e^{\alpha+p^2/2MkT}-1}$$

$$=\frac{V}{h^3}(2\pi MkT)^{\frac{3}{2}}\cdot F(\alpha), \qquad (2)$$

where

$$F(\alpha)=\frac{2}{\pi^{\frac{1}{2}}}\int_0^\infty \frac{z^{\frac{1}{2}}dz}{e^{\alpha+z}-1}$$

$$=e^{-\alpha}+\frac{e^{-2\alpha}}{2^{\frac{3}{2}}}+\frac{e^{-3\alpha}}{3^{\frac{3}{2}}}+\frac{e^{-4\alpha}}{4^{\frac{3}{2}}}+\cdots.$$

Here α must be positive; otherwise some of the N_i would be negative, which, of course, is not admissible. Now $F(\alpha)$ is a monotonously decreasing function, its maximum value being $F(0)=2.612$. Therefore, no solution $\alpha(T)$ of Eq. (2) can be found for

$$N>\frac{V(2\pi MkT)^{\frac{3}{2}}}{h^3}2.612,$$

i.e., for $$T<T_0,$$

where $$T_0=\left(\frac{N}{2.612V}\right)^{\frac{2}{3}}\frac{h^2}{2\pi Mk}. \qquad (3)$$

For M=mass of the He-atom and for a molar volume of 27.6 cm³ one obtains $T_0=3.13°$K.

Equation (1), however, obviously always has a solution $\alpha(T)$, for each temperature T, with positive N_i. The difference of the behavior of Eqs. (1) and (2) comes from the neglect caused by the substitution of the sum by an integral. This can easily be seen if one takes, e.g., the discontinuous energy values of a cubic volume of linear dimension L (periodic boundary conditions):

$$\epsilon_{k,l,m}=(k^2+l^2+m^2)(h^2/2ML^2)$$

$$k,l,m=0,\pm1,\pm2,\cdots$$

and substitutes them into (1):

$$N=\sum_{k,l,m}\frac{1}{e^{\sigma(k^2+l^2+m^2)+\alpha}-1}, \qquad (1a)$$

where $$\sigma=\frac{h^2}{2L^2MkT}=\frac{1.90\pi}{N^{\frac{2}{3}}}\cdot\frac{T_0}{T}.$$

Here the right-hand side is again a monotonously decreasing function of α, which, however, is not bounded by a finite limit when $\alpha\to0$. Consequently Eq. (1′) has, for each temperature T, a solution $\alpha(T)$.[5] The difference between (1a) and (2) is due to the fact that in (2) the lowest state $k=l=m=0$ (or $p=0$), incorrectly acquires the statistical weight zero, and therefore it becomes entirely suppressed. It is just this *one* term in the sum (1a) which is decisive for the behavior in the limit $\alpha\to0$, *viz.*, at lowest temperatures. It will therefore be sufficient to treat this lowest term of the sum separately; the other terms may, without detriment, be replaced by an integral.

We collect the states with $k^2+l^2+m^2\leqq\rho^2$ where the value of ρ may be chosen in such a way, that the difference between the integral from ρ to ∞ and the sum from ρ to ∞ is negligible (say $\rho\sim10$). The mean energy of these states is

$$\bar{\epsilon}_0=\frac{3}{10}\frac{\rho^2h^2}{ML^2}=\frac{3}{5}\epsilon \quad\text{where}\quad \epsilon=\frac{h^2\rho^2}{2ML^2}$$

and their number is given by

$$g_0=(4\pi/3)\rho^3.$$

[5] From this fact Uhlenbeck[2] was led to the conclusion that no condensation phenomenon should exist.

Then we may write, in a better approximation than (2),

$$N=\frac{g_0}{e^{\beta\tilde{\epsilon}_0+\alpha}-1}+\frac{4\pi V}{h^3}\int_{\rho h/L}^{\infty}\frac{p^2dp}{e^{\alpha+p^2/2MkT}-1}$$

or with $\alpha'=\beta\tilde{\epsilon}_0+\alpha=\frac{3}{5}\beta\epsilon+\alpha$:

$$1=\frac{g_0/N}{e^{\alpha'}-1}+\left(\frac{T}{T_0}\right)^{\frac{3}{2}}\frac{2}{\pi^{\frac{1}{2}}\cdot F(0)}\times\int_{\beta\epsilon}^{\infty}\frac{z^{\frac{1}{2}}dz}{e^{z+\alpha'-3/5\cdot\beta\epsilon}-1}. \quad (4)$$

The first term of the right-hand side must not be negative, since it gives the number of atoms in the cell of the g_0 lowest states. Therefore in any case $\alpha'>0$. It may be remarked that the quantity $\beta\epsilon=(h^2\rho^2/2ML^2kT)$ is always a very small number $[=(5.97/N^{\frac{2}{3}})\rho^2(T_0/T)]$.

Now we have to distinguish two entirely different cases:

(1) $T>T_0$. In this case α' comes out to be a number of the order of magnitude 1. Consequently, the first term of the right-hand side of (4) is of the order $1/N$ and may be neglected compared with the second term. In the latter we may neglect $\beta\epsilon\ll\alpha'$ and obtain

$$(T_0/T)^{\frac{3}{2}}=[F(\alpha')/F(0)] \quad \text{for } T>T_0, \quad (4a)$$

which is equivalent to (2).

(2) $T<T_0$. The second term of the right-hand side of (4) is in any case smaller than $(T/T_0)^{\frac{3}{2}}$. In particular for $T<T_0$, it is smaller than 1. Thus the first term is required to make up for the rest; consequently, since the numerator of this term is very small, its denominator must be equally small, *viz.*, of the order $1/N$, this means $\alpha'\sim 1/N$; in the second term α' may now be neglected compared with $\beta\epsilon\sim 1/N^{\frac{2}{3}}$ and we can write for (4):

$$1=\frac{g_0}{N\alpha'}+(T/T_0)^{\frac{3}{2}}\frac{2}{\pi^{\frac{1}{2}}F(0)}\int_{\beta\epsilon}^{\infty}\frac{z^{\frac{1}{2}}dz}{e^{z-3/5\cdot\beta\epsilon}-1} \quad (4b)$$

$$=\frac{g_0}{N\alpha'}+(T/T_0)^{\frac{3}{2}}[1-0.173(\beta\epsilon)^{\frac{1}{2}}+\cdots]$$

or

$$\alpha'=\frac{g_0}{N}[1-(T/T_0)^{\frac{3}{2}}(1-0.173(\beta\epsilon)^{\frac{1}{2}}+\cdots)]^{-1}, \quad (5)$$

valid as long as

$$N[1-(T/T_0)^{\frac{3}{2}}]\gg 1 \quad \text{or for} \quad T_0-T\gg\frac{2}{3N}T_0,$$

i.e., practically for all temperatures $T<T_0$.

The number of particles in the lowest cell becomes

$$N_0=N[1-(T/T_0)^{\frac{3}{2}}(1-0.173(\beta\epsilon)^{\frac{1}{2}}+\cdots)]$$

$$=N[1-(T/T_0)^{\frac{3}{2}}]+0.422\rho N^{\frac{1}{2}}T/T_0+\cdots.$$

We see that in first approximation this expression is independent of ρ, i.e., independent of $g_0=(4\pi/3)\rho^3$, the number of states we have collected in the lowest cell. Therefore, $N(1-(T/T_0)^{\frac{3}{2}})$ atoms, i.e., *a finite fraction of all atoms* will be assembled in the *one* lowest state. In the spherical shell of radius ρ and thickness $d\rho$ around the lowest state, there will be only $0.422\ N^{\frac{1}{2}}T/T_0d\rho$ molecules.

Accordingly, the distribution among the lowest quantum states, characterized by the quantum numbers k, l, m (see (1a)), may be written:

$$N_{k,l,m}=\frac{N[1-(T/T_0)^{\frac{3}{2}}]}{1+(k^2+l^2+m^2)N^{\frac{1}{2}}29.8T_0/T[1-(T/T_0)^{\frac{3}{2}}]}.$$

If N is finite, the function $\alpha(T)$ is, of course, an analytic function. However, for increasing values of N, the third derivative of α, near $T=T_0$, becomes greater and greater. In the limit $N\to\infty$, $N/V=$constant, the function $\alpha(T)$ has a discontinuous second derivative. It consists of two branches which do not cohere analytically. One gets $\alpha=0$ for $T\leqslant T_0$, while for $T\geqslant T_0$ the function $\alpha(T)$ is given by the inversion of (4a). Taking $N=$Avogadro's number, and $V=$molecular volume, we may, for $T<T_0$, write simply:

$$N_0=N[1-(T/T_0)^{\frac{3}{2}}]$$
(number of atoms in the lowest state) (6a)

$$N(E)dE=(2\pi V/h^3)(2M)^{\frac{3}{2}}(E^{\frac{1}{2}}dE/(e^{\beta E}-1)).$$
(number of atoms of the interval $(E, E+dE)$ where $E>0$) (6b)

Therefrom we obtain the energy U per molecular

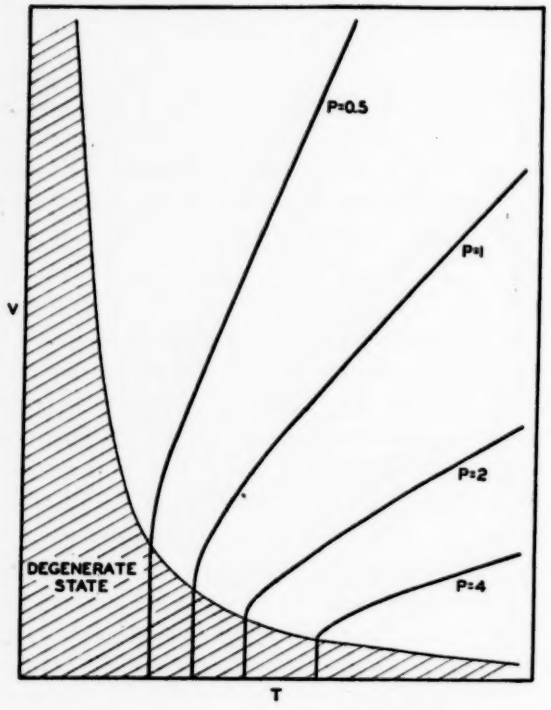

FIG. 1. Isobars of a Bose-Einstein gas.

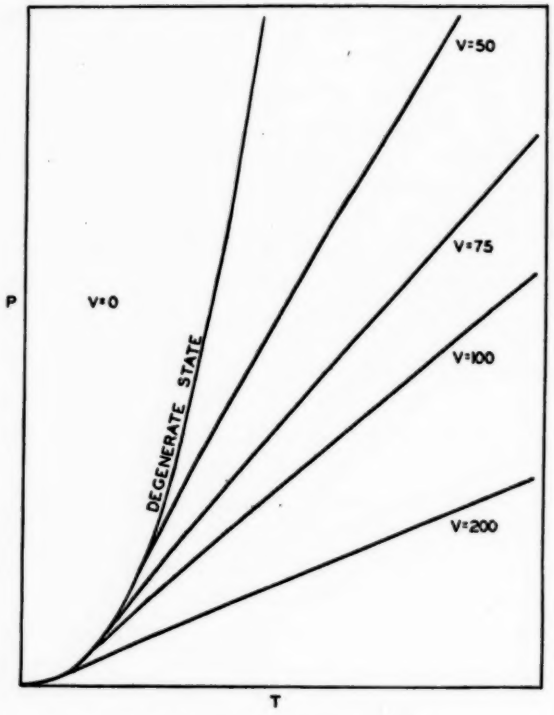

FIG. 2. Isochores of a Bose-Einstein gas.

volume and for $T<T_0$:

$$U_-=\int_0^\infty EN(E)dE=\frac{3}{2}RT\left(\frac{T}{T_0}\right)^{\frac{3}{2}}\frac{\zeta(2.5)}{\zeta(1.5)}$$

$$=\frac{3}{2}0.514RT\left(\frac{T}{T_0}\right)^{\frac{3}{2}}=\frac{3}{2}0.514RCT^{5/2}V, \quad (7a)$$

where $C=\dfrac{2.612}{N}\left(\dfrac{2\pi Mk}{h^2}\right)^{\frac{3}{2}}\approx\dfrac{1}{153}$ for He

and $$\zeta(k)=\frac{1}{(k-1)!}\int_0^\infty\frac{x^{k-1}dx}{e^x-1}.$$

denotes the Riemann zeta-function ($\zeta(\frac{3}{2})=2.612$; $\zeta(\frac{5}{2})=1.341$).

For $T\geqslant T_0$, Einstein has previously given the semi-convergent expansion:

$$U_+=\frac{3RT}{2}\{1-0.462(T_0/T)^{\frac{3}{2}}$$

$$-0.0225(T_0/T)^3-0.0114(T_0/T)^{9/2}-\cdots\}. \quad (7b)$$

The two branches U_+ and U_- are continuous at $T=T_0$ with a continuous tangent, but the second derivative is discontinuous. Thus the specific heat c_v has a *break* at $T=T_0$ (see Fig. 4):

$$c_{v-}=\frac{15}{4}0.514R(T/T_0)^{\frac{3}{2}} \quad (8a)$$

$$c_{v+}=\tfrac{3}{2}R\{1+0.231(T_0/T)^{\frac{3}{2}}$$

$$+0.045(T_0/T)^3+0.040(T_0/T)^{9/2}+\cdots\}. \quad (8b)$$

The free energy

$$F=-T\int_0^T\frac{U}{T^2}dT$$

is found to be:

$$F_-=-0.514RT(T/T_0)^{3/2}=-0.514CRT^{5/2}V \quad (9a)$$

$$F_+=-\tfrac{3}{2}RT\{\ln(T/T_0)+0.308(T_0/T)^{3/2}$$

$$+0.0075(T_0/T)^3+0.0025(T_0/T)^{9/2}+\cdots\}, \quad (9b)$$

from which one gets the pressure $p=-\partial F/\partial V$

$$p_-=0.514CRT^{5/2} \quad (10a)$$

$$p_+=\frac{RT}{V}\left[1-\frac{0.462}{CVT^{\frac{3}{2}}}-\frac{0.0225}{(CVT^{\frac{3}{2}})^2}\right.$$

$$\left.-\frac{0.0114}{(CVT^{\frac{3}{2}})^3}-\cdots\right]. \quad (10b)$$

This is in agreement with the Virial theorem: $pV=\frac{2}{3}U$.

§2. Condensation in Momentum Space

Attention has already been directed to the remarkable result (10a), that for $T<T_0$ the pressure is independent of density, just as in the case of a van der Waals gas in the transition region below the critical point. However, this independence of density is quite accidental for the Bose-Einstein gas,[6] and it is somewhat misleading to interpret this behavior in analogy with a van der Waals gas by saying that for $T<T_0$ a fraction of the molecules condenses into a state of zero volume and does not contribute to the density.[4] Indeed, in the transition region of the van der Waals gas, a fraction of the molecules is condensed into a state of considerably smaller volume and, since it is separated by gravitation from the other molecules, it does not contribute to the density of the gaseous phase. At this point the analogy fails completely, or is rather to be found in a more remote sense. The supernumerary molecules of a degenerate Bose-Einstein gas which we have denoted by N_0 and which indeed may be considered as belonging to a particular phase, do not, of course, disappear mysteriously from space; they do contribute to the density as any other molecules. They do *not contribute, however, to the pressure*, since their kinetic energy (and momentum), is zero. If one likes analogies, one may say that there is actually a *condensation, but only in momentum space*, and not in ordinary space, i.e., an equilibrium of two phases, *one* containing the molecules N_0 of momentum zero and occupying in the space of momenta, a zero volume; and *another one* showing a distribution over all momenta similar to that which is realized for $T>T_0$. In ordinary space, however, no separation of phases is to be noticed.

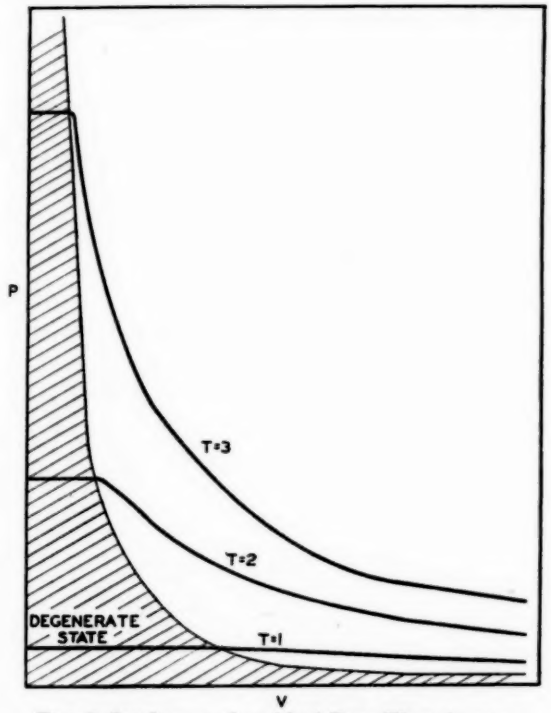

FIG. 3. Isotherms of an ideal Bose-Einstein gas.

In a certain respect the molecules of the condensed phase having the momentum zero, show also a characteristic peculiarity as to their behavior in ordinary space. Since their wave functions are constant over the whole volume, they are particularly inappropriate for forming wave packets of the size of molecular dimensions by superposition of neighboring wave functions (i.e., wave functions of comparable small energy, so

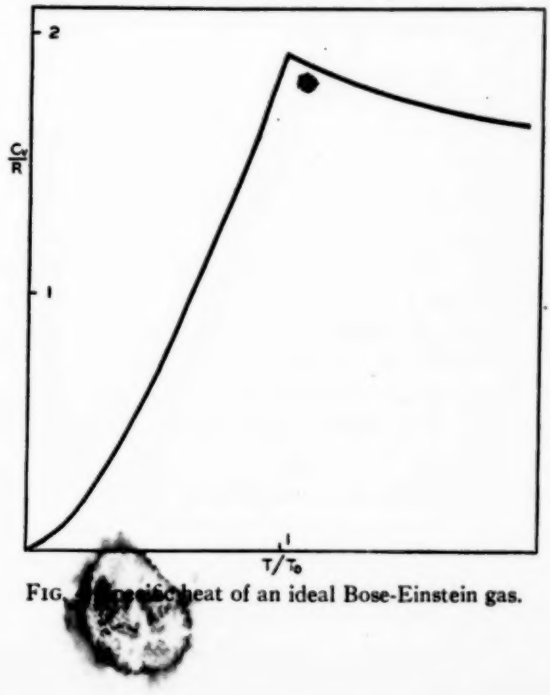

FIG. [illegible] specific heat of an ideal Bose-Einstein gas.

[6] This independence of density holds only for an "ideal" Bose-Einstein gas, but is no longer realized as soon as a molecular interaction is taken into account, for instance by adding a term $U_0(V)$ to the energy, representing a mean van der Waals field smeared over the total volume accessible to the molecules. In this case the condensation for constant volume proceeds as before and the specific heat c_v is exactly the same as given by (8). The pressure, however, will no longer be independent of V. It will be given by

$$p_-=0.514\ CRT^{5/2}-\partial U_0/\partial V$$

and for constant pressure the gas will in this case no longer condense to the volume zero.

that the energy distribution of the whole remains unchanged). These molecules therefore will represent a peculiar omnipresence in the total volume at their disposal. Accordingly it will not be allowable to treat their motion in external fields (pressure gradient, etc.), on the lines of Ehrenfest's theorem (approximate validity of classical mechanics for small wave packets), which theorem has been the general basis of the usual corpuscular treatment of the transport phenomena. Here we have before us just the opposite limiting case, namely, wave packets of a very *small extension* in the space of momenta, but which in ordinary space are spread over a region comparable with the extension of the inhomogeneities of macroscopic fields.

The quantum dynamics of this limiting case has been investigated very little. Only one interesting special case has so far been discussed in connection with superconductivity. It has been shown[7] that the macroscopic description given for this phenomenon leads to a relation between electric current and magnetic field, which is identical with that which would be valid in an enormous diamagnetic atom of the dimensions of the metal. Now it is well known that diamagnetism cannot be explained on the basis of classical dynamics of electrons. The currents in a diamagnetic atom are not given by progressive de Broglie waves (or progressive wave packets), as in the usual treatment of the ordinary conductivity phenomena. These currents are represented rather by standing waves; they come from that term in the expression for the electric current which contains the vector potential of the magnetic field, and their structure is intimately connected with the spatial extension *of the quantum states* (the diamagnetic susceptibility is known to be proportional to $\langle r^2 \rangle_{Av}$). For the superconductivity in particular it can be shown that the macroscopic phenomena may be interpreted very simply by the assumption of a peculiar coupling in the space of momenta, producing, below a certain critical temperature, a fixation of momentum, as if there were something like a condensed phase in the space of momenta. Thus far, however, it has not yet been possible to base that assumption on a molecular model by the general theory of electrons in metals.

[7] F. London and H. London, Physica **2**, 341 (1935); F. London, Proc. Roy. Soc. **A152**, 24 (1935).

Now the degenerate Bose-Einstein gas provides a good example of a molecular model for such a condensed state in space of momenta, such as seems needed for the superconductive state. Though this fact cannot, of course, be applied to explain superconductivity, as electrons do not obey Bose statistics, it is remarkable on the other hand that the transport properties (viscosity,[8] thermal conductivity[9]), of liquid helium, when passing the λ-point actually change in a very conspicuous manner; thus one speaks of a "superfluidity" and of a super-heat-conductivity."

§3. Superfluidity

We have already emphasized that one might certainly not be justified in formally applying the ordinary corpuscular theory[10] of transport phenomena of a Bose-Einstein gas to this case of degeneracy where the *mean de Broglie wave-length of the particles is comparable with the macroscopic dimensions of the whole system*. On the present occasion, however, it may perhaps be permissible to discuss, with all due reserve, the question of what would be the effect if, in Sommerfeld's theory of conductivity,[11] the Fermi statistics are formally replaced by the Bose-Einstein statistics, and in particular, what would happen at the transition to the degenerate state. One may presume that our attempt will prove as much justified as the classical treatment of diamagnetism on the basis of the Larmor theorem.

In order to fix our ideas and to have a definite, though highly idealized, model, we may depict liquid helium as a metal in which ions and electrons are replaced by particles of the *same* kind, namely both by He atoms. Each He atom may move in a self-consistent (van der Waals) field formed by the other atoms. The states of this system might perhaps, to a certain approxima-

[8] E. F. Burton, Nature **135**, 265 (1935); Wilhelm, Misener and Clark, Proc. Roy. Soc. (London) **A151**, 342 (1935); E. F. Burton, Nature **142**, 72 (1938); P. Kapitza, Nature **141**, 74 (1938); J. F. Allen and A. D. Misener, Nature **141**, 75 (1938); B. V. Rollin, VII. Congrès Internat. du Froid **1**, 187 (1936); Kikoin and Lasarew, Nature **142**, 912 (1938); J. G. Daunt, and K. Mendelssohn, Nature **141**, 911 (1938); **142**, 475 (1938).

[9] W. H. Keesom and A. P. Keesom, Physica **2**, 557 (1935); B. V. Rollin, Physica **3**, 296 (1936); W. H. Keesom and A. P. Keesom, Physica **3**, 359 (1936); J. F. Allen, R. Peierls and M. Z. Uddin, Nature **140**, 62 (1937).

[10] G. E. Uhlenbeck and E. A. Uehling, Phys. Rev. **43**, 552 (1933).

[11] A. Sommerfeld, Zeits. f. Physik **47**, 1 (1928).

tion, be divided into two classes: one, being more of the *Bloch-type* (progressive modulated Schrödinger waves), corresponding to the *electronic states* of the metal and representing a transport of matter; the other one, more of the *Debye-type* (quantized acoustical or elastic waves), corresponding to the *vibrations of the ionic lattice*. In this picture the *fluidity* of the liquid would correspond to the electric *conductivity* of the electrons in a metal; the friction would be due to the dissipation of progressive Bloch waves by inelastic reflection on the Debye waves. Let us finally make the assumption that for these inelastic collisions there exists a mean free path l which is independent of the velocity v for small values of v.

We do not want to insist here on the details of this conception, which is certainly open to much criticism, though perhaps it may prove useful for describing some properties of liquids in general. At any rate it is desirable to know what would be the result if in the theory of metals the Fermi statistics are formally replaced by Bose statistics, even if one does not think of a possible connection with liquid helium, and even if it is more than doubtful whether one is allowed to apply the ordinary collision theory to this limiting case in which the mean value of the de Broglie wavelength is comparable with the macroscopic dimensions of the whole system.

The well-known formula for the electric conductivity $\sigma \sim \langle l/v \rangle_{Av}$ would yield the value $\sigma = \infty$ in the case where a finite fraction of all particles has zero velocity. But actually one has to apply a special consideration for the slowest particles. A particle of mass M and of initial velocity v_0, chosen in the direction of the field F, needs the time τ to traverse the mean free path l:

$$l = v_0\tau + (1/2M)F\tau^2,$$

or
$$\tau = (M/F)[(v_0^2 + 2lF/M)^{\frac{1}{2}} - v_0].$$

The mean change of velocity due to the action of F during this time τ will therefore be given by

$$\langle \delta v_0 \rangle_{Av} = \frac{1}{2M} F\tau = \tfrac{1}{2}[(v_0^2 + 2lF/M)^{\frac{1}{2}} - v_0]. \quad (11)$$

In general (particularly in the case of Fermi statistics, and also in the case $T > T_0$ for the Bose statistics), one is accustomed and entitled to disregard the few particles with $Mv_0^2/2 \lesssim lF$ and one may write instead of (11):

$$\langle \delta v_0 \rangle_{Av} = \frac{l}{2Mv_0} \cdot F \quad \text{for} \quad Mv_0^2/2 \gg lF, \quad (11a)$$

which gives the above-mentioned formula for σ.

In the case of the Bose degeneracy ($T < T_0$), however, we are not allowed to neglect the *finite fraction* of particles with $v_0 = 0$. For these we obtain

$$\langle \delta v_0 \rangle_{Av} = (lF/2M)^{\frac{1}{2}} \quad \text{for} \quad Mv_0^2/2 \ll l \cdot F. \quad (11b)$$

This contribution is, for ordinary values of F, enormously greater than (11a). One may say the "conductivity," *viz.*, the *fluidity* (defined as derivative of the current with respect to F), becomes *infinite* for $F \to 0$, and this abruptly, as soon as $T < T_0$. For a fixed value of F, the current will be proportional to N_0/N, i.e., proportional to the fraction of atoms of velocity zero.

The particles N_0, having the energy zero, will not appreciably contribute to the transport of energy, and therefore, we should not expect a particularly great increase of the *heat conductivity* when passing to the degenerate state.

§4. Thermomechanical Effect

But there is another mechanism which may produce a transfer of heat.[12] The van der Waals forces between the walls and the He atoms are much stronger than those between the He atoms themselves. In a layer, L, of perhaps 10A, or 100A, or even greater thickness along the walls, the van der Waals field will be appreciably stronger than in the interior, I, of the liquid. In this layer the concentration of the degenerate atoms will therefore be much greater than in the interior, the entropy in the layer, S_L, will be much smaller than in the interior, S_I. Thus we have a situation quite analogous to a thermocouple: namely, two different conductors, L and I, in conducting contact.

If q mols of helium pass at a temperature T from I to L they will go into a state of greater order, and the heat

$$Q = T(S_I - S_L)q \quad (12)$$

will be set free. This will occur in a *reversible manner*, quite as in the case of the Peltier effect. Relating all transfer of energy or matter to unit

[12] H. [illegible] do, Nature **142**, 612 (1938).

of time and to unit of cross section, we may write

$$Q = T(S_I - S_L)J, \tag{12a}$$

where J is the current density. The "Peltier coefficient" is accordingly given by

$$\Pi_{I,\,L} = T(S_I - S_L). \tag{13}$$

Assuming for S the expression for an ideal degenerate Bose gas, which is given by (7) (9)

$$S_- = \frac{U_- - F_-}{T} = (5/2)0.514CRT^{5/2}V, \tag{14}$$

we obtain for the Peltier coefficient

$$\Pi_{I,\,L} = (5/2)0.514CRT^{5/2}(V_I - V_L). \tag{15}$$

Now the arrangement for measuring heat conductivity can simply be considered as a thermocouple consisting of the two "metals" L and I. If we heat at one point of the wall and cool at another one, in such a way that the temperature at the two points may be kept constant at T_1 and T_2 respectively, we may produce a "thermoforce" Φ given, according to W. Thomson's well-known thermodynamic relation, by the formula

$$\Phi = \int_{T_1}^{T_2} \frac{\Pi_{I,\,L}}{T} dT = \int_{T_1}^{T_2} (S_I - S_L) dT$$

or with (14):

$$\Phi = (T_2 - T_1)(5/2)0.514CRT_1^{\frac{3}{2}}(V_I - V_L). \tag{16}$$

This "thermoforce" will produce very great circulation of matter, since the "internal resistance" of the thermoelement is extraordinarily small, and, therefore, the consumption of heat for maintaining even the smallest temperature differences will also be very great. This process, therefore, will appear like an enormous conduction of heat; the "conductivity," however, will depend very strongly on the gradient of temperature. For the lowest temperatures, the helium atoms will everywhere be almost completely degenerate, in the layer as well as in the interior of the liquid. There will be no difference of entropy between the different parts of the liquid, the "thermoforce" will disappear—as it must according to Nernst's Law—and there will no longer be a production of circulation in the liquid. Only ordinary heat conductivity (chiefly by the Debye vibrations), will remain. Nevertheless the great fluidity, since it is proportional to the number N_0 of degenerate atoms, will persist down to the lowest temperatures; only the driving thermoforce and the Peltier heat will disappear.

All this is in qualitative agreement with the experiments, particularly with those of Keesom and Saris,[13] and of Kürti and Simon,[13] on the heat conductivity of He at the lowest temperatures, which below 0.6°K, has been found to become "normal" again, i.e., small and independent of the gradient of temperature. In fact, this transition to normal heat conductivity occurs in just that region where the thermal anomaly of the specific heat, connected with the λ-point, ceases and passes over into an ordinary Debye T^3-law for the specific heat.

This mechanism of reversible transformation of heat into mechanical energy gives a very simple explanation also for the so-called "fountain phenomenon" observed by Allen and Jones,[14] and interprets it as a pump driven by a thermoelement. It may be remarked that according to (12) the flow of matter in the capillary layer has to have the opposite direction to the transfer of heat, if the entropy in the capillary layer is smaller than in the normal liquid. In fact this connection between the directions of the flow of heat and of the flow of matter has actually been observed in the experiment of Allen and Jones. The same has been found in the Knudsen manometer experiment of H. London.[12]

The idea that the transport phenomena of He II might be reversible processes has first been discussed by L. Tisza.[15] The thermodynamic relation (12) has recently been given by H. London.[12] Though it might appear that the logical connection between the facts will not be qualitatively very different from the one we have sketched here, it is obvious that the theoretical basis given thus far is not to be considered more than a quite rough and preliminary approach to the problem of liquid helium, limited chiefly by the lack of a satisfactory molecular theory of liquids.

The author wishes to express his thanks to Professor P. Gross for his hospitality in the Chemistry Department of Duke University, and for his kind interest in this work.

[13] W. H. Keesom, A. D. Keesom, and B. F. Saris, Physica **5**, 281 (1938); N. Kürti and F. Simon, Nature **142**, 207 (1938).
[14] J. F. Allen and H. Jones, Nature **141**, 243 (1938).
[15] L. Tisza, Nature **141**, 913 (1938).

A Method for Treating Large Perturbations

H. A. BETHE
Cornell University, Ithaca, New York
(Received October 7, 1938)

A method is developed for treating large perturbations to which the ordinary Schrödinger perturbation theory is not applicable. The method can be used for all Hamiltonian matrices which are sufficiently regular. The condition is that the matrix elements in any given diagonal should change slowly from one row to the next (cf. §2). It is not necessary that the Hamiltonian be expressible as an explicit function of coordinates and momenta. The wave function is expanded in the usual way in "unperturbed eigenfunctions," $\psi = \Sigma_n c(n)\psi_n$. The expansion coefficients $c(n)$ are then calculated by a method similar to the well-known WKB method; only we deal with a difference rather than a differential equation (§3). It is found that $c(n)$ behaves, as a function of n, either exponentially or oscillatory or exponentially with alternating sign (§4), the first two types of behavior being similar to the ordinary WKB, while the last type is peculiar to difference equations. Several examples are given to illustrate the method, *viz.* some perturbed oscillator problems (§3, 5*A*, *B*) and the Matthieu functions (§5*C*). After some generalizations (§6, 7), the method is applied to fast electrons in metals (§8) whose wave functions are determined in momentum space. It is shown that a plane wave is an insufficient approximation even for high energy.

§1. INTRODUCTION

THE Schrödinger perturbation theory is one of the most widely used methods in quantum theory. It is only applicable to small perturbations, and in general only the first approximation which gives a nonvanishing result is useful: If this approximation does not yield a sufficiently accurate result, the convergence is slow and it is preferable to discard perturbation methods altogether. On the other hand, the desirability of a theory of large perturbations is obvious, not only for the quantitative results obtainable in this way, but also because perturbation calculations are often more easily interpreted than direct integrations.

A good example is the theory of electrons in metals. When Sommerfeld first developed this theory,[1,2] he assumed the electrons to be free. It is well known how successful this picture has been in explaining the phenomena of metallic conduction, thermoelectric effects, etc. However, Sommerfeld realized already in his first paper that the assumption of free electrons had to be justified by quantum theory. While in a qualitative way a complete justification could be given,[3] it was found impossible[4] to calculate quantitatively the eigenfunctions of metal electrons by applying perturbation theory to free electrons because the perturbation is much too large. On the other hand, the opposite picture of slightly perturbed atomic wave functions also led to unsatisfactory results. The solution was given by the direct method of Wigner and Seitz,[5] but this method is primarily applicable to valence electrons and becomes very cumbersome for highly excited states. It might be thought that for these states the free electron approximation was satisfactory, but actually this is not the case even for the highest energies (cf. §7) so that a new method is needed.

In this paper, I shall give a method for treating large perturbations. It is valid for all perturbations which are sufficiently large so that the ordinary perturbation theory is no longer applicable, provided that the perturbation fulfills certain regularity requirements (§2). We shall apply this method to several perturbed oscillator problems, to the Matthieu functions and to fast metal electrons.

§2. NOTATION

We shall assume that the matrix elements of the Hamiltonian are given in some reference system. The diagonal elements

$$H_{nn} = W(n) \tag{1}$$

are ordered according to their size. The difference

$$\Delta(n+\tfrac{1}{2}) = W(n+1) - W(n) \tag{2}$$

[1] Sommerfeld, Naturwiss. **15**, 825 (1927).
[2] Sommerfeld, Zeits. f. Physik **47**, 1 (1928).
[3] Bloch, Zeits. f. Physik **52**, 555 (1929).
[4] Sommerfeld and Bethe, *Handbuch der Physik, 24/2*, 333 (1933).
[5] Wigner and Seitz, Phys. Rev. **43**, 804 (1933).

will be called the "spacing." For the nondiagonal elements we introduce the notation

$$H_{n,\ n+k}=H_{n+k,\ n}=-A_k(n+\tfrac{1}{2}k). \quad (3)$$

The minus sign is convenient for the following. When we use A's with negative subscript, we define

$$A_{-k}(n)=A_k(n). \quad (3a)$$

(For simplicity, we assume all matrix elements to be real which is true in most problems if the diagonal elements W_n are all different, i.e., if the original system is nondegenerate).

We make the following assumptions:

I. Regularity assumption

The matrix elements A and the spacings Δ shall be smooth functions of n, i.e.,

$$|A_k(n+1)-A_k(n)|\ll A_k(n), \quad \text{(all } k\text{)} \quad (4)$$

$$|\Delta(n+1)-\Delta(n)|\ll\Delta(n). \quad (5)$$

II. Magnitude assumption

For a given n, at least one of the A's shall be large compared with the corresponding spacing, i.e.,

$$|A_k(n+\tfrac{1}{2}k)|\gg W(n+k)-W(n) \quad \text{(for at least one } k\text{)}. \quad (6)$$

This is the condition for nonapplicability of the ordinary Schrödinger perturbation theory.

The "unperturbed" wave function corresponding to $W(n)$ will be called ψ_n, and the actual wave function is expanded:

$$\psi=\sum_n c(n)\psi_n. \quad (7)$$

Then the coefficients c must fulfill the relation

$$c(n)(W(n)-\epsilon)=\sum_{k=-\infty}^{\infty}{}' c(n+k)A_k(n+\tfrac{1}{2}k), \quad (8)$$

where ϵ is the eigenvalue. The function c of n will often be referred to as "wave function." The prime at the sum over k means omission of $k=0$. Our problem is the determination of ϵ and of the c's. The method will be similar to the Wentzel-Kramers-Brillouin method for the approximate solution of differential equations. It will first be developed in connection with a special example.

§3. Oscillator with Constant Nondiagonal Perturbation

As the simplest problem, we take an oscillator so that the unperturbed eigenvalues are equally spaced. Choosing proper units, we may put

$$W(n)=n. \quad (9)$$

n takes on all positive half-integer values, *viz.* $\frac{1}{2}$, $\frac{3}{2}$, 5/2, etc. The perturbation shall have matrix elements only next to the main diagonal, and these shall all be equal:

$$A_1(n)=\tfrac{1}{2}A \qquad (\text{all } n;\ A\gg 1) \quad (10)$$

$$A_k(n)=0. \qquad (k\neq 1,\ \text{all } n) \quad (10a)$$

Then (8) becomes

$$c(n)(n-\epsilon)=\tfrac{1}{2}A[c(n-1)+c(n+1)], \quad (11)$$

which must be fulfilled for all n's. The "boundary conditions" on the function $c(n)$ are

$$c(-\tfrac{1}{2})=0, \qquad c(\infty)=0. \quad (12)$$

The system of Eq. (11) has some similarity to the equations for the vibration of a linear chain of atoms coupled by elastic forces between neighbors. The term $c(n)$ is to be identified with the amplitude of the nth atom, $\epsilon=\omega^2$ where ω is the frequency of the vibration, and A is connected with the elastic force. The only difference is that the variable n in the factor $n-\epsilon$ is replaced by the constant A. The solution of the linear chain problem is a simple wave for $c(n)$,

$$c(n)=a\sin kn. \quad (13)$$

In our case, the solution cannot be so simple because of the occurrence of the variable factor n in (11). However, since n varies "slowly" (spacing between two eigenvalues small compared with A), we can try a solution similar to (13) but with slowly variable amplitude a and wave number k, just as in the familiar WKB method. We put

$$c(n)=a(n)\sin\chi(n) \quad (14)$$

$$\chi(n)=\sum_{\nu=0}^{n-\frac{1}{2}}\varphi(\nu) \quad (14a)$$

and assume a and φ to be slowly variable.

We insert (14) into (11), replace $\sin \chi(n\pm 1)$ by $\sin \chi(n) \cos \varphi(n\pm\frac{1}{2}) \pm \cos \chi(n) \sin \varphi(n\pm\frac{1}{2})$ and put the factors of $\sin \chi(n)$ and $\cos \chi(n)$ separately equal to zero. Then we expand $a(n\pm 1)$ and $\varphi(n\pm\frac{1}{2})$ in Taylor series, and neglect all terms containing more than one first derivative or any higher derivative. Then the factor of $\sin \chi(n)$ gives

$$\cos \varphi(n) = (n-\epsilon)/A \tag{15}$$

and the factor of $\cos \chi$:

$$a(n) = \gamma[\sin \varphi(n)]^{-\frac{1}{2}}, \tag{15a}$$

where γ is a constant. Neglecting consistently the second derivative of φ, we may replace the sum (14a) by an integral, and obtain the final result

$$c(n) = \gamma[\sin \varphi(n)]^{-\frac{1}{2}} \sin \int_{-\frac{1}{2}}^{n} \varphi(\nu)d\nu. \tag{16}$$

This formula, and Eq. (15), resemble closely the WKB formulas. By choosing the sin rather than the cos in (16), we fulfill the boundary condition (12) at $n=-\frac{1}{2}$.

We discuss Eq. (15) for positive A. (For negative A, we need only insert a factor $(-1)^n$ in (16) and then replace A by $|A|$ in (15).) Let us first assume

$$-A < \epsilon < A. \tag{17}$$

Then $\varphi(0)$ is real, and can, without loss of generality, be taken between 0 and π. $\varphi(n)$ will decrease with increasing n. That means that $c(n)$ is represented by a wave of increasing wave-length. Simultaneously, the amplitude changes according to (15a). This continues up to

$$n_2 = \epsilon + A, \tag{17a}$$

where φ reaches the value zero. For $n > n_2$, Eq. (15) no longer has a real solution for φ. As in the ordinary WKB, the oscillatory wave function (16) then goes over into an exponential solution. In order to avoid infinities for large n, it is necessary that this solution be exponentially *decreasing*. Then it has the form

$$c(n) = \tfrac{1}{2}\gamma[\sinh f(n)]^{-\frac{1}{2}} \exp\left[-\int_{n_2}^{n} f(\nu)d\nu\right] \tag{18}$$

for $n > n_2$

with

$$\cosh f(n) = (n-\epsilon)/A. \tag{18a}$$

In order that (16) join on to the exponentially *decreasing* function (18), it is necessary that

$$\chi(n_2) = \int_{-\frac{1}{2}}^{n_2} \varphi(\nu)d\nu = (k+\tfrac{3}{4})\pi, \tag{19}$$

where k is an integer (quantum number). (19) is the quantum condition which determines the eigenvalues.

If we use Eq. (15) for φ and carry out the integration in (19), we find

$$(A^2-\epsilon'^2)^{\frac{1}{2}} + \epsilon' \arccos(-\epsilon'/A) = (k+\tfrac{3}{4})\pi, \tag{20}$$

from which

$$\epsilon' = \epsilon + \tfrac{1}{2} \tag{20a}$$

can be calculated as a function of k. For the lowest eigenvalues ($\epsilon' = -A+\delta$, with $\delta \ll A$), we have the explicit formula

$$\epsilon' = -A + \tfrac{1}{2}(3\pi)^{\frac{1}{3}} A^{\frac{1}{3}} (k+\tfrac{3}{4})^{\frac{2}{3}} + 0(A^{-\frac{1}{3}}). \tag{21}$$

The spacing of the low eigenvalues is thus

$$D \approx d\epsilon/dk \approx 1.49(A/k)^{\frac{1}{3}}, \tag{21a}$$

i.e., larger than for the unperturbed oscillator for which it is unity. The eigenfunction of the lowest state ($k=0$) is composed of oscillator functions up to

$$n_2 \approx 1.84 A^{\frac{1}{3}}; \tag{21b}$$

the coefficients $c(n)$ are largest slightly below n_2. The function $c(n)$ has k nodes for the kth quantum state.

Quite generally, we find by differentiation of (20) for the spacing of the energy levels:

$$D = \pi/\arccos(-\epsilon'/A). \tag{22}$$

E.g., for $\epsilon' = -\frac{1}{2}A\sqrt{3}$, the spacing is $D=6$; for $\epsilon'=0$, $D=2$, and for $\epsilon'=A$, D reaches the value 1, i.e., the same as for the unperturbed oscillator. In this case, φ goes from π to 0 when n goes from 0 to $n_2=2A$, and the "wave function" $a(n)$ is large at both limits, 0 and n_2.

When ϵ becomes larger than A, Eq. (15) has no real solution at $n=0$ and up to

$$n_1 = \epsilon - A. \tag{22a}$$

Then a solution of exponential character will

also hold for $n<n_1$. The solution which satisfies the first boundary condition (12) is

$$c(n)=(-1)^{n-\frac{1}{2}}\gamma'[\sinh f(n)]^{-\frac{1}{2}}\sinh\int_{-\frac{1}{2}}^{n}f(\nu)d\nu, \quad (23)$$

where

$$\cosh f(n)=-(n-\epsilon)/A \quad (23a)$$

and

$$\gamma'\approx\gamma\exp\left(-\int_{-\frac{1}{2}}^{n_1}f(\nu)d\nu\right). \quad (23b)$$

The $c(n)$'s for $n<n_1$ alternate in sign and increase almost exponentially up to n_1. Then the wave function c begins to oscillate with continuously increasing wave-length. In the range from n_1 to n_2, c is given by

$$\begin{aligned}c(n)&=(-1)^{n-\frac{1}{2}}\gamma[\sin\varphi(n)]^{-\frac{1}{2}}\\&\quad\times\cos\left\{\int_{n_1}^{n}(\pi-\varphi(\nu))d\nu-\tfrac{1}{4}\pi\right\}\\&=\gamma[\sin\varphi(n)]^{-\frac{1}{2}}\\&\quad\times\cos\left(\pi n_1+\int_{n_1}^{n}\varphi(\nu)d\nu-\tfrac{1}{4}\pi\right)\end{aligned} \quad (23c)$$

with φ as defined in (15). The amplitude of the oscillations is largest near the two ends of the oscillatory region, n_1 and n_2. Beyond n_2, there is again an exponential decrease of $c(n)$ but now with *constant* sign. The quantum condition becomes

$$\pi n_1+\int_{n_1}^{n_2}\varphi(n)dn-\tfrac{1}{4}\pi=(k+\tfrac{1}{4})\pi. \quad (24)$$

The integral can be carried out and gives $\pi(\epsilon-n_1)$. Therefore

$$\epsilon=k+\tfrac{1}{2}. \quad (24a)$$

This means that for $\epsilon>A$ the eigenvalues are exactly the same as for the unperturbed oscillator while the eigen*functions* are, of course, different. The effect of the perturbation on the spectrum is to spread the A eigenvalues below A over the region between $-A$ and $+A$ instead of from 0 to A.

It is interesting to compare our "WKB method for difference equations" with the ordinary WKB for differential equations. The quantity $T=\epsilon+A-n$ is the analog to the kinetic energy $T=E-V$ in the ordinary theory. When T becomes negative, we obtain a solution of exponential type in both cases. However, when T is positive, there is a striking difference: For the differential equation, $T>0$ always means an oscillating wave function; for the difference equation, this is only true as long as T is smaller than $2A$ while for large T we have an exponential type again. The connection with the high frequency limit of crystal vibrations is obvious.

§4. General Method for "Three-Diagonal" Matrices

We shall now consider a general Hamiltonian matrix containing elements in the principal diagonal and in the two diagonals adjoining, i.e., a "three-diagonal" matrix. The elements shall be subject to the regularity and magnitude conditions of §2, and to the convergence condition

$$|A_1(n)|<\tfrac{1}{2}W(n). \quad (n \text{ large}) \quad (25)$$

Otherwise they shall be unrestricted. The lowest eigenvalue shall be arbitrarily labeled $n=1$. Then, in general, the boundary conditions will be

$$c(0)=0, \quad c(\infty)=0; \quad (26)$$

the first of them follows from the absence of a matrix element $H_{0,1}$.

We define

$$\Phi(n)=[W(n)-\epsilon]/2A_1(n) \quad (27)$$

and distinguish three cases:

$$\text{(a)} \quad -1<\Phi(n)<+1 \quad (28a)$$

$$\text{(b)} \quad \Phi(n)>+1 \quad (28b)$$

$$\text{(c)} \quad \Phi(n)<-1. \quad (28c)$$

The complete domain of n, from 0 to ∞, will contain at least two different regions (of type a and b), but in general more (see below).

The form of the wave function $c(n)$ in the different regions is as follows:

(a) In regions of type (a), the wave function is oscillating

$$c(n)=\gamma[A_1(n)\sin\varphi(n)]^{-\frac{1}{2}}\cos\int^{n}\varphi(\nu)d\nu. \quad (29a)$$

(b) In regions of type (b) we have an exponential behavior (decreasing or increasing)

$$c(n)=\gamma[A_1(n)\sinh f(n)]^{-\frac{1}{2}} \times\exp\pm\int^n f(\nu)d\nu. \quad (29b)$$

Linear combinations, such as cosh or sinh, are, of course, also possible.

(c) In regions of type (c) the solution is also of exponential type but with the sign of the c's alternating

$$c(n)=\gamma(-1)^n[A_1(n)\sinh f(n)]^{-\frac{1}{2}} \times\exp\pm\int^n f(\nu)d\nu. \quad (29c)$$

The lower limits of the integrals are determined by smoothly joining the solutions in the various regions which will be discussed below.

The variable "wave number" φ (or f) is connected with the quasi-"potential" Φ by

$$\cos\varphi(n)=\Phi(n) \quad \text{(case a)} \quad (30a)$$

$$\cosh f(n)=\Phi(n) \quad \text{(case b)} \quad (30b)$$

$$\cosh f(n)=-\Phi(n). \quad \text{(case c)} \quad (30c)$$

The solutions (29) (30) can be verified by direct insertion into the equation determining the $c(n)$, *viz.*

$$(W(n)-\epsilon)c(n)=A_1(n+\tfrac{1}{2})c(n+1) +A_1(n-\tfrac{1}{2})c(n-1). \quad (31)$$

The quantities neglected are second derivatives, or products of two first derivatives, of φ and $\log A_1$. These are, by hypothesis, small of the second order.

We discuss now the joining of the solutions in the various regions. Because of the convergence condition[6] (25), there will *always* be a region of type (b) for large n. In this region, the solution is exponential; because of the boundary conditions, it must be exponentially *decreasing;* we may put

$$c(n)=\gamma[A_1(n)\sinh f(n)]^{-\frac{1}{2}} \times\exp-\int_{n_2}^n f(\nu)d\nu, \quad (32b)$$

where n_2 is defined by $\Phi(n_2)=1$.

Adjoining the exponential region, there must be[7] an oscillating region (type a) towards smaller n. The phase of the wave function in this region is determined by the condition that it must join to the exponentially decreasing solution (32b). In the usual way,[7] we find for the solution in region a:

$$c(n)=2\gamma[A_1(n)\sin\varphi(n)]^{-\frac{1}{2}} \times\cos\left(\int_{n_2}^n\varphi(\nu)d\nu+\tfrac{1}{4}\pi\right). \quad (32a)$$

An arbitrary multiple of π may, of course, be added to the argument of the cos.

Regarding the behavior of $\Phi(n)$ for $n<n_2$, there are various possibilities:

I. Φ may remain between the limits -1 and $+1$ down to $n=0$. Then the solution (32a) holds everywhere between 0 and n_2, and the boundary condition $c(0)=0$ demands that c behaves as

$$c(n)\sim\sin\int_0^n\varphi(\nu)d\nu \quad (33a)$$

(leaving out the first factors). Identification of (33a) and (32a) leads to the quantum condition

$$\int_0^{n_2}\varphi(n)dn=(k+\tfrac{3}{4})\pi, \quad (33)$$

where k is any non-negative integer. If the boundary condition at $n=0$ were $dc/dn=0$, we should take the cos instead of the sin in (33a), and instead of (33) we should have

$$\int_0^{n_2}\varphi(n)dn=(k+\tfrac{1}{4})\pi. \quad (33b)$$

II. Φ may again become larger than $+1$ for small n. This can be the case only for very small

[6] If (25) is *not* fulfilled, we shall have $\Phi(n)<1$ for large n, and therefore the wave function $c(n)$ will continue to oscillate up to $n=\infty$ whatever ϵ. In this case, we have a continuous rather than a discrete spectrum, which is beyond the scope of this paper.

[7] If the exponential solution (32b) extended down to $n=0$, the boundary condition $c(0)=0$ (cf. 26) could not be fulfilled.

ϵ, more accurately for

$$\epsilon < W(0) - 2A_1(0). \tag{34}$$

Such low eigenvalues exist only for certain problems, e.g. not for the problem of §3. If (34) is fulfilled there will be another "type a" region between 0 and n_1, say. In order to fulfill the boundary condition $c(0)=0$, the solution in this region must be

$$c(n) = \gamma'[A_1 \sinh f(n)]^{-\frac{1}{2}} \sinh \int_0^n f(\nu)d\nu. \tag{34a}$$

If n_1 is not too small, this behaves very nearly as $\exp(+\cdots)$ near n_1. Therefore the solution in the oscillatory region must be

$$c(n) = \gamma' \exp \int_0^{n_1} f(\nu)d\nu \,[A_1 \sin \varphi(n)]^{-\frac{1}{2}} \times \cos\left\{\int_{n_1}^n \varphi(\nu)d\nu - \tfrac{1}{4}\pi\right\}. \tag{34b}$$

Identification of (32a) and (34a) yields the quantum condition

$$\int_{n_1}^{n_2} \varphi(n)dn = (k+\tfrac{1}{2})\pi. \tag{35}$$

Since $\varphi(n_1)=\varphi(n_2)=0$, this can also be written (integration by parts!)

$$\int_{n_2}^{n_1} n\,d\varphi = (k+\tfrac{1}{2})\pi. \tag{35a}$$

The relation between γ and γ' is:

$$2\gamma = \gamma' \exp \int_0^{n_1} \varphi(\nu)d\nu. \tag{35b}$$

The quantum condition (35) differs from (33) by having $\frac{1}{2}$ instead of $\frac{3}{4}$, it is identical with the condition of the ordinary WKB.

III. $\Phi(n)$ becomes smaller than -1 for small n. This will *always* be the case for sufficiently large ϵ, more accurately for

$$\epsilon > W(0) + 2A_1(0). \tag{36}$$

Then there is a "type c" region between 0 and n_1 where $\Phi(n_1) = -1$. The solution for $0 < n < n_1$ is

$$c(n) = (-1)^{n-1}\gamma'[A_1 \sinh f(n)]^{-\frac{1}{2}} \times \sinh \int_0^n f(\nu)d\nu. \tag{36a}$$

In order to join the oscillatory solution to (36a), we have also to put the factor $(-1)^{n-1}$ in evidence in it. This can be done by writing at the same time $\pi-\varphi$ rather than φ in the integrand of the cos which has also the advantage that $\pi-\varphi$ is very small for n slightly greater than n_1. It can easily be seen that the correct continuation of (36a) is

$$c(n) = (-1)^{n-1}\gamma' \times \exp\left(\int_0^{n_1} f(\nu)d\nu\right)[A_1 \sin \varphi(n)]^{-\frac{1}{2}} \times \cos\left[\int_{n_1}^n (\pi - \varphi(\nu))d\nu - \tfrac{1}{4}\pi\right]. \tag{37}$$

This can now be rewritten

$$c(n) = 2\gamma[A_1 \sin \varphi(n)]^{-\frac{1}{2}} \times \cos\left[\pi(n-1) - \int_{n_1}^n \{\pi - \varphi(\nu)\}d\nu + \tfrac{1}{4}\pi\right] = \cdots \cos\left[\pi(n_1 - \tfrac{3}{4}) + \int_{n_1}^n \varphi(\nu)d\nu\right], \tag{37a}$$

where γ is related to γ' by (35b). Identification of (37a) with (32a) gives the quantum condition

$$\pi n_1 + \int_{n_1}^{n_2} \varphi(n)dn = (k+1)\pi. \tag{38a}$$

Integrating by parts, and remembering that $\varphi(n_2)=0$, $\varphi(n_1)=\pi$, we obtain

$$\int_0^{\pi} n\,d\varphi = (k+1)\pi, \tag{38}$$

where k is again the number of changes of sign in the wave function $c(n)$, from $n=1$ to ∞. Eq. (38) corresponds to "integral quantum numbers" instead of the half-integral ones in case II or the quarter integers of case I.

IV. It may happen that there are *more than three different "regions"* between 0 and ∞. E.g., $\Phi(n)$ may be >1 for $n>n_2$, then between -1 and 1 for $n_1<n<n_2$, then <-1 for $n_0<n<n_1$, then again between -1 and $+1$, etc. This corresponds to the existence of two (or more) potential holes in ordinary wave mechanics. Each oscillatory region (type a) is equivalent to one

potential hole. The wave function is large in the oscillatory regions (sometimes not in all of them). The quantum condition is easily derivable in each particular case.

As regards the occurrence of the various cases, we have already mentioned that case III will always occur for very high ϵ. For the lowest eigenvalues we shall in general have case I if $A_1(0)$ is nonvanishing, case II, if $A_1(0)=0$ and $\epsilon<W_1(0)$, and case III, if $A_1(0)=0$ and $\epsilon>W_1(0)$.

§5. Examples

A first and very simple example was treated in §3. In this section, we shall give some examples which can be solved rigorously by the ordinary methods.

A.—We take an oscillator so that $W(n)=n=\frac{1}{2}, \frac{3}{2}, \cdots$ and put

$$A_1(n)=Bn^{\frac{1}{2}}. \tag{39}$$

Then

$$\Phi(n)=(n-\epsilon)/2Bn^{\frac{1}{2}}. \tag{40}$$

The critical points n_1 and n_2 are obtained by putting $\Phi(n)=\pm 1$. This gives

$$n_{1,2}=[(B^2+\epsilon)^{\frac{1}{2}}\pm B]^2. \tag{41}$$

By definition, $n_2>n_1$. By inserting $n_{1,2}$ into (40), it is easily seen that

$$\Phi(n_2)=+1 \quad \text{(as always)} \tag{41a}$$

$$\Phi(n_1)=\begin{cases}+1 & \text{for } \epsilon<0\\ -1 & \text{for } \epsilon>0.\end{cases} \tag{41b}$$

Therefore we have case II of §4 for negative ϵ, and case III for positive ϵ. For negative ϵ, the quantum condition (35) gives immediately

$$\begin{aligned}\frac{1}{\pi}\int_{n_1}^{n_2}\varphi(\nu)d\nu=\frac{1}{\pi}\Big\{&n \arccos\frac{n-\epsilon}{2Bn^{\frac{1}{2}}}\\ &+(\epsilon+B^2)\arcsin\frac{n-\epsilon-2B^2}{2B(B^2+\epsilon)^{\frac{1}{2}}}\\ &-[B^2n-\tfrac{1}{4}(\epsilon-n)^2]^{\frac{1}{2}}\Big\}_{n_1}^{n_2}\end{aligned} \tag{42}$$

$$\equiv\epsilon+B^2=k+\tfrac{1}{2}.$$

For positive ϵ, we must take into account that in our present case the first value of n is $\frac{1}{2}$ whereas in (38) it was 1. Therefore $k+1$ on the r.h.s. of (38) should be replaced by $k+\frac{1}{2}$. Carrying out the integral in (38), we obtain again

$$\epsilon+B^2=k+\tfrac{1}{2}.$$

Thus the eigenvalues are the same as for an oscillator of frequency 1, only shifted downwards by $-B^2$.

This very simple result is easily understood. The unperturbed problem with the eigenvalues $W(n)=n=\frac{1}{2}, \frac{3}{2}$, etc. is

$$\tfrac{1}{2}d^2\psi/dx^2+(W-\tfrac{1}{2}x^2)\psi=0. \tag{43}$$

The perturbation matrix $H_{n,n+1}=B(n+\frac{1}{2})^{\frac{1}{2}}$ is equivalent to a perturbing potential $\sqrt{2}Bx$; therefore the perturbed problem is

$$\begin{aligned}&\tfrac{1}{2}d^2\psi/dx^2+(\epsilon-\tfrac{1}{2}x^2-\sqrt{2}Bx)\psi\\ &\quad\equiv\tfrac{1}{2}d^2\psi/dx^2+[\epsilon+B^2-\tfrac{1}{2}(x+\sqrt{2}B)^2]\psi=0,\end{aligned} \tag{43a}$$

i.e., an oscillator equal to that in (43) except for a shift of the equilibrium position to $-\sqrt{2}B$ and replacement of W by $\epsilon+B^2$.

B.—Another simple example is an oscillator with changed frequency. Let the "perturbed problem" be

$$\tfrac{1}{2}d^2\psi/d\dot{x}^2+(\epsilon-\tfrac{1}{2}\beta^2x^2)\psi=0. \tag{44}$$

Then the perturbed eigenvalues are

$$\epsilon=\beta(k+\tfrac{1}{2}). \tag{44a}$$

If we consider (43) as the unperturbed problem, the perturbing potential is

$$V=\tfrac{1}{2}(\beta^2-1)x^2, \tag{45}$$

its matrix elements

$$V_{nn}=\tfrac{1}{2}(\beta^2-1)n \tag{45a}$$

$$V_{n,n+2}=\tfrac{1}{4}(\beta^2-1)(n+\tfrac{1}{2})^{\frac{1}{2}}(n+\tfrac{3}{2})^{\frac{1}{2}}\approx\tfrac{1}{4}(\beta^2-1)(n+1). \tag{45b}$$

Therefore the matrix of the "perturbed" Hamiltonian (cf. §2):

$$W(n)=\tfrac{1}{2}(\beta^2+1)n \tag{46}$$

$$A_2(n)=\tfrac{1}{4}(\beta^2-1)n \tag{46a}$$

$$A_k(n)=0 \qquad \text{for } k\neq 2. \tag{46b}$$

These matrix elements obey the convergence condition (25).

Obviously, the correct eigenfunction contains only every second n (either $n=2m+\frac{1}{2}$ or $n=2m-\frac{1}{2}$, with m integer). We have case III whatever ϵ, and, for $\beta>1$

$$n_1=\epsilon/\beta^2, \quad n_2=\epsilon. \tag{47}$$

The quantum condition is

$$\tfrac{1}{2}\pi(n_1-\tfrac{1}{2})+\tfrac{1}{2}\int_{n_1}^{n_2}\varphi dn=k\pi, \tag{47a}$$

where the factor $\frac{1}{2}$ arises from the fact that only those $c(n)$ occur for which $n-\frac{1}{2}$ is even. Evaluation of the integral gives

$$\epsilon=\beta(2k+\tfrac{1}{2}) \tag{48}$$

as in (44a) (*even* states!).

C. Matthieu functions.—We consider the differential equation

$$d^2\psi/dx^2+(\epsilon+A\cos x)\psi=0 \tag{49}$$

(simple periodic potential). Those solutions which are periodic with the same period as the potential, *viz.* 2π, are the Matthieu functions.

As "unperturbed problem" we take the equation of the free particle

$$d^2\psi/dx^2+W\psi=0. \tag{50}$$

The (normalized) solutions of period 2π are

$$\psi_n=(2\pi)^{-\frac{1}{2}}e^{inx}. \tag{50a}$$

where n is an integer; the corresponding eigenvalue

$$W(n)=n^2. \tag{50b}$$

The matrix elements of the perturbing potential $H=A\cos x$ are

$$A_1(n+\tfrac{1}{2})=H_{n,\,n+1}=\tfrac{1}{2}A \quad \text{(all } n\text{)} \tag{51}$$

$$A_k(n)=0 \qquad (k\neq 1, \text{ all } n). \tag{51a}$$

A is supposed to be large compared with unity.

Since $W(n)$ and $A(n)$ are both even in n, the solutions of the "perturbed" problem fall into two classes, those even and those odd in n, i.e.,

$$c(-n)=\pm c(n). \tag{51b}$$

The "even" solutions behave like

$$\cos\int_0^n \varphi d\nu \quad \text{or} \quad \cosh\int_0^n f d\nu$$

the "odd" ones like sin or sinh. This symmetry condition replaces the boundary condition $c(0)=0$ (Cf. 26).

For $-A<\epsilon<A$, the behavior of Φ is according to class I (cf. §4). If the solution is odd (boundary condition $c(0)=0$), the quantum condition (33) applies; for even solutions, the boundary condition is effectively $dc/dn=0$ so that we have the quantum condition (33b). For $\epsilon>A$, we have an example of class III of §4, so that (38) holds. We have

$$n_1=(\epsilon-A)^{\frac{1}{2}}, \quad n_2=(\epsilon+A)^{\frac{1}{2}}. \tag{51c}$$

The eigenfunction $c(n)$ is large between n_1 and n_2, i.e., in an interval of length

$$n_2-n_1\approx A\epsilon^{-\frac{1}{2}}. \tag{51d}$$

For large $\epsilon(>A^2)$, this becomes <1; then ordinary perturbation theory holds.

Remembering (27), (30a), the quantum condition is generally

$$\int_0^{\varphi_0} n d\varphi=\int_0^{\varphi_0}(A\cos\varphi+\epsilon)^{\frac{1}{2}}d\varphi=(k+\delta)\pi, \tag{52}$$

where for $\epsilon<A$ ("class I")

$$\varphi_0=\varphi(0)=\arccos(-\epsilon/A), \quad \delta=\pm\tfrac{1}{4} \tag{52a}$$

(the + sign for even, the − sign for odd states). For $\epsilon>A$ (class III)

$$\varphi_0=\pi, \quad \delta=0. \tag{52b}$$

The condition (52) follows also from a direct application of the ordinary WKB method to the original Eq. (49), with φ replaced by x. Also the values of δ following from such a treatment are the same as in (52a, b). The integral in (52) can be evaluated in terms of complete elliptic integrals.

§6. Extension to Matrices with Many Nondiagonal Elements

When more than three diagonals of the Hamiltonian matrix contain nonvanishing elements, it is convenient to write the wave function $c(n)$ in exponential form (rather than cos and sin), *viz.*

$$c(n)=\text{R. P. } a(n)\exp\int^n f(\nu)d\nu. \tag{53}$$

(R. P. = Real part of). From the fundamental equation (8) we find

$$2\sum_{k=1}^{\infty} A_k\cosh[kf(n)]=W(n)-\epsilon \tag{54}$$

and

$$a(n)=\gamma\left[\sum_{k=1}^{\infty} kA_k\sinh kf(n)\right]^{-\frac{1}{2}}, \tag{55}$$

where γ is a constant, in general complex.

Equation (54) serves for the determination of $f(n)$; Eq. (55) gives $a(n)$ in terms of $f(n)$. The relations are obtained by expanding f and A in a Taylor series about n; then the first term (containing no derivatives) yields (54), the second (containing one first derivative) gives (55) and the higher terms are neglected, as usual. We may simplify (54) (55) by introducing the abbreviation

$$F(f,n)\equiv 2\sum_{k=1}^{\infty} A_k(n)\cosh kf. \tag{56}$$

Then

$$F(f,n)=W(n)-\epsilon \tag{56a}$$

$$a(n)=\gamma(\partial F/\partial f)^{-\frac{1}{2}}. \tag{56b}$$

While the formal solution (53)–(55) is very similar to the case of three-diagonal matrices, the

explicit solution of (54) is somewhat more complicated. For three-diagonal Hamiltonians, we have seen that the solution is either simply exponential (f real) or oscillatory (f purely imaginary) or exponential with alternating sign (f=real part $+\pi i$). In the general case, more general complex values of f will occur.

A. We begin with the simple problem

$$A_1(n)=A_2(n)=\tfrac{1}{2}A \quad \text{(all } n\text{)} \tag{57}$$

$$A_k(n)=0. \qquad (k>2) \tag{57a}$$

Then

$$F\equiv A\,(\cosh f+\cosh 2f)=W(n)-\epsilon. \tag{57b}$$

For very large n, the solution f is obviously real so that we have a simple exponential solution (which must, of course, be decreasing). For a certain $n=n_2$ for which

$$W(n_2)=\epsilon+2A, \tag{58}$$

we obtain $f=0$. For $n<n_2$, f becomes purely imaginary,

$$f=i\varphi, \quad \cos\varphi+\cos 2\varphi=(W(n)-\epsilon)/A. \tag{58a}$$

The left-hand side of (58a) decreases with increasing φ; however, it does not continue to decrease until $\varphi=\pi$ but reaches a minimum for $\varphi=104.47°$, $\cos\varphi=-\frac{1}{4}$. This minimum has the value $-9/8$. Therefore the oscillatory region of the wave function stops at n_1 where

$$W(n_1)=\epsilon-(9/8)A. \tag{58b}$$

In contrast to the case of a three-diagonal Hamiltonian, the oscillatory region is not symmetrical about the point $W(n)=\epsilon$ but extends farther towards large than towards small n.

For $n<n_1$, the solution of (57b) becomes complex. Putting

$$f=\psi+i\varphi \tag{59}$$

$$\Phi=(-W(n)+\epsilon)/A, \tag{59a}$$

(57b) gives:

$$\cos\varphi\cosh\psi+\cos 2\varphi\cosh 2\psi=-\Phi \tag{59b}$$

$$\sin\varphi\sinh\psi+\sin 2\varphi\sinh 2\psi=0. \tag{59c}$$

(59b) represents the real, (59c) the imaginary part. From the latter equation,

$$\cos\varphi=-1/(4\cosh\psi). \tag{60}$$

This means that φ tends towards $\pi/2$ as ψ increases (i.e., as n decreases). When $\varphi=\pi/2$, the sign of $c(n+2)$ is opposite to the sign of $c(n)$. Inserting (60) into (59b), we find

$$2\cosh\psi=(\tfrac{1}{2}\Phi)^{\frac{1}{2}}+(\tfrac{1}{2}\Phi+1)^{\frac{1}{2}}. \tag{60a}$$

For $\Phi=9/8$, this gives $\psi=0$; for larger Φ, ψ increases.

B. As a second example, we take

$$A_1(n)=A_2(n)=\cdots=A_m(n)=A \tag{61}$$

$$A_k(n)=0. \quad (k>m) \tag{61a}$$

Then

$$F+A=A\frac{\sinh(m+\frac{1}{2})f}{\sinh\frac{1}{2}f}=W(n)+A-\epsilon. \tag{61b}$$

For large n, we shall again have real f; for $n=n_2$, with

$$W(n_2)=\epsilon+2mA, \tag{62}$$

we obtain $f=0$; for $n<n_2$, f is purely imaginary,

$$f=i\varphi, \quad \frac{\sin(m+\frac{1}{2})\varphi}{\sin\frac{1}{2}\varphi}=\frac{W(n)+A-\epsilon}{A}. \tag{62a}$$

If m is large, the left-hand side will reach its minimum already for very small φ, *viz.* for

$$\varphi=4.49/(m+\tfrac{1}{2}). \tag{62b}$$

This value of φ will be reached for $n=n_1$ where n_1 is determined by

$$W(n_1)=\epsilon-A-0.217A(2m+1). \tag{63}$$

From (62), (63) we see that the oscillatory region is even more unsymmetrical than in the previous example. For $n<n_1$, the solution of (61b) is complex; putting

$$f=(x+iy)/(m+\tfrac{1}{2}), \tag{63a}$$

we have

$$\frac{\sinh x\cos y}{x}=\frac{\cosh x\sin y}{y}=\frac{W(n)+A-\epsilon}{A(2m+1)} \tag{63b}$$

and therefore

$$\tanh x/x=\tan y/y \tag{64}$$

$$\frac{\sinh x}{(y^2\tanh^2 x+x^2)^{\frac{1}{2}}}=\frac{-W(n)-A+\epsilon}{A(2m+1)}. \tag{64a}$$

As n decreases from n_1 to smaller values, y

decreases from 4.49 (where $\tan y = y$) to π, and x increases according to (64a). When the right-hand side of (64a) is very large, the left-hand side becomes simply $\sinh x/x$.

C. As a third case, we take

$$A_k(n) = A_0 e^{-\alpha^2 k^2} \tag{65}$$

with α very small. Then

$$F \approx \int_{-\infty}^{\infty} A_0 e^{-\alpha^2 k^2 + kf} dk = \pi^{\frac{1}{2}} A_0 \alpha^{-1} e^{f^2/4\alpha^2}. \tag{65a}$$

f is real for $n > n_2$ where

$$W(n_2) = \epsilon + \pi^{\frac{1}{2}} A_0/\alpha. \tag{66}$$

The oscillatory region (f purely imaginary) extends down to n_1 where

$$W(n_1) = \epsilon. \tag{66a}$$

For $n < n_1$, we have from (65a)

$$f = 2\alpha \left[\pi i + \log \frac{\alpha(\epsilon - W(n))}{\pi^{\frac{1}{2}} A_0} \right]^{\frac{1}{2}}. \tag{67}$$

As n decreases from n_1, f is at first almost purely imaginary, then the real part becomes steadily larger. Real and imaginary part are equal when

$$W(n) = \epsilon - \pi^{\frac{1}{2}} A_0/\alpha, \tag{67a}$$

then

$$f = (2\pi)^{\frac{1}{2}} \alpha (1+i). \tag{67b}$$

The two last examples have in common that f remains in general small, in all the three domains (real, imaginary and complex). This means that consecutive c's are not very different from one another, the "wave function" $c(n)$ is rather smooth. This is a direct consequence of the presence of *many* large nondiagonal elements in the Hamiltonian, all of the same sign. If $A_k(n)$ is large for $k < m$ and becomes smaller thereafter, the wave function $c(n)$ will change considerably only if n increases by an amount of the order m. This "smoothness" of the wave function is obviously very helpful.

§7. Case of Divergent $\Sigma_k A_k(n)$

Thus far, we have only discussed cases in which

$$\sum_{k=1}^{\infty} A_k \text{ converges.} \tag{68}$$

If this is not the case, which happens frequently in practice, the determination of the upper limit n_2 of the oscillatory region is somewhat more complicated. Inside the oscillatory region itself, there is no additional difficulty, since

$$\Sigma A_k(n) \cos \varphi$$

converges, for $\varphi \neq 0$, provided only $A_k \to 0$ for $k \to \infty$. But near n_2 the representation (53) of $c(n)$ breaks down (this is always true but in general not important) so that the fundamental Eq. (8) must be solved directly. A simplification is possible because c changes only slowly with n (cf. §6, end); therefore the sum in (8) may be replaced by an integral:

$$(W(n) - \epsilon) c(n) = \int_{-\infty}^{\infty} A_k(n + \tfrac{1}{2}k) c(n+k) dk. \tag{69}$$

In order to solve this integral equation near the transition from oscillatory to exponential behavior, we use again the analogy to the ordinary WKB method as applied to second-order differential equations. Near the transition point n_2, the solution of any differential equation is approximately similar to that of the simple equation

$$d^2\omega/dx^2 + x\omega = 0. \tag{70}$$

The solution

$$\omega = (\pi^{\frac{1}{2}}/3)[J_{-1/3}(\tfrac{2}{3}x^{\frac{3}{2}}) + J_{1/3}(\tfrac{2}{3}x^{\frac{3}{2}})] \tag{70a}$$

has been tabulated by Kramers.[8] We shall now assume that our function c is also similar to ω which assumption will actually lead to a solution. We put

$$c(n) \sim \omega((n - n_2)/\lambda), \tag{71}$$

where the transition point n_2 and the scale factor λ must yet be determined.

One equation for this purpose is obtained by inserting (71) in (69) and putting $n = n_2$. Another is derived by differentiating (69) with respect to n:

$$\frac{dW}{dn} c(n) + (W(n) - \epsilon) \frac{dc(n)}{dn} = \int_{-\infty}^{\infty} \left\{ A_k(n + \tfrac{1}{2}k) \frac{dc(n+k)}{dn} + \frac{dA_k(n+\tfrac{1}{2}k)}{dn} c(n+k) \right\} dk. \tag{71a}$$

[8] Kramers, Zeits. f. Physik **39**, 828 (1926).

Here we can consistently neglect the term with dA/dn and replace $A_k(n+\frac{1}{2}k)$ by $A_k(n)$, since A is supposed to change slowly. Then we multiply (71a) by $c(n_2)$, (69) by $(dc/dn)_{n_2}$, subtract and put $n=n_2$. Remembering (71), we get

$$\left(\frac{dW}{dn}\right)_{n_2}\omega^2(0)=\int_{-\infty}^{\infty}A_k(n_2)\times\left\{\omega(0)\frac{d\omega(k/\lambda)}{dk}-\frac{d\omega}{dk}(0)\omega(k/\lambda)\right\}dk$$

$$=\int_{-\infty}^{\infty}A_{\lambda x}(n_2)\left\{\omega(0)\frac{d\omega}{dx}(x)-\frac{d\omega}{dx}(0)\omega(x)\right\}dx. \quad (72)$$

This equation determines λ. Both dW/dn and $A_k(n)$ do not change very rapidly with n; thus the left-hand side is to all purposes a given constant while on the right-hand side the scale factor λ in $A_{\lambda x}$ is the only quantity on which the integral depends.

It is essential for the applicability of this method that $A_0 \gg dW/dn$ (cf. §2). In order that (72) be fulfilled, it is then necessary that A_λ is already comparatively small because (72) means, in effect, that an average of the A's, over k from 0 to several times λ, is of the order of the spacing dW/dn. Thus λ is of the order of *that* value of k at which the A_k's begin decreasing rapidly. When λ is determined, n_2 follows from (69):

$$[W(n_2)-\epsilon]\omega(0)=\lambda\int A_{\lambda x}(n_2)\omega(x)dx. \quad (72a)$$

The determination of n_1 offers no difficulties in principle; we must seek only the minimum of the convergent sum

$$F=\sum_k A_k\cos k\varphi \quad (73)$$

as a function of φ; then

$$W(n_1)=\epsilon+F_{\min}. \quad (73a)$$

In the "exponential" regions, however, the behavior of $c(n)$ is different from previous cases: (53) breaks down since $\Sigma_k A_k \cosh kf$ would diverge hopelessly for real f. $c(n)$ is no longer determined by the neighboring c's but by the c's in the oscillatory region; therefore the decrease of c with increasing n is primarily due to the decrease of A_k with k.

§8. Three-Dimensional Case: Electrons in Metals

We consider a simple lattice of atoms with screened Coulomb potential. We require the solutions of the wave equation which have the periodicity of the lattice.[9] As the unperturbed problem, we take free electrons. Then the unperturbed wave functions are

$$\psi(\mathbf{K})=\Omega_0^{-\frac{1}{2}}\exp i\mathbf{K}\cdot\mathbf{r}, \quad (74)$$

where $\mathbf{K}$ is 2π times a vector in the reciprocal lattice[4] and Ω_0 is the volume per atom. The number of states in an interval of $\mathbf{K}$ is

$$(\Omega_0/8\pi^3)dK_x dK_y dK_z, \quad (74a)$$

the unperturbed eigenvalue is proportional to K^2, we put

$$W(\mathbf{K})=K^2, \quad (74b)$$

so that W and ϵ are the energy values times $2m/\hbar^2$. The matrix elements of the perturbing potential (screened Coulomb potential) are then

$$A_k=\frac{2m}{\hbar^2}V_k=\frac{2m}{\hbar^2}Ze^2\int\frac{1}{r}\exp(-\alpha r+i\mathbf{k}\cdot\mathbf{r})\frac{d\tau}{\Omega_0}$$

$$A_k=\frac{8\pi Z}{a_0\Omega_0}\frac{1}{\alpha^2+k^2} \quad (75)$$

with $a_0=\hbar^2/me^2$ the Bohr radius. $A_{\mathbf{k}}(\mathbf{K})$ is independent of $\mathbf{K}$.

Then the fundamental equation (8) becomes

$$(K^2-\epsilon)c(\mathbf{K})=\sum_{\mathbf{k}}A_{\mathbf{k}}c(\mathbf{K}+\mathbf{k})=\frac{Z}{\pi^2 a_0}\int\frac{d\mathbf{k}}{\alpha^2+k^2}c(\mathbf{K}+\mathbf{k}). \quad (76)$$

The problem has spherical symmetry and can therefore be separated in polar coordinates:

$$c(\mathbf{K})=\psi(K)P_{lm}(\vartheta\varphi) \quad (77)$$

with P_{lm} a spherical harmonic. For high energy, most of the eigenfunctions have large l (those of small l can easily be treated separately). Instead of considering all different m's for a given l, we

[9] It is equally easy to find a wave function $\psi(\mathbf{r})$ which is multiplied by $e^{i\alpha}$, $e^{i\beta}$, $e^{i\gamma}$ when $\mathbf{r}$ is increased by the identity period in one of the three principal directions of the crystal. (α, β, γ are then the components of the reduced wave vector.)

may as well consider the functions $m=0$ for different directions of the polar axis z. This is more closely related to the problem of electron diffraction which is the most interesting problem involving electronic states of high energy in solids. In electron diffraction, we have the electron incident along a certain direction z; therefore we want, if possible, a wave function which is large only for K's pointing approximately in that direction.

For large l, we may write[10]

$$P_l(\vartheta)=J_0((l+\tfrac{1}{2})\vartheta) \tag{78}$$

(J_0 the Bessel function). With (77), (78) and the addition theorem of spherical harmonics, Eq. (76) becomes

$$(K^2-\epsilon)\psi(K)=(Z/\pi^2a_0)\times\int\frac{2\pi K'^2dK'\sin\vartheta d\vartheta}{(K-K')^2+\alpha^2+2KK'(1-\cos\vartheta)}\times\psi(K')J_0(l'\vartheta). \tag{79}$$

Here we have put $\mathbf{K}'=\mathbf{K}+\mathbf{k}$ and $l'=l+\frac{1}{2}$. Now we introduce

$$z^2=2KK'(1-\cos\vartheta)\approx K^2\vartheta^2 \tag{79a}$$

$$b^2=(K-K')^2+\alpha^2. \tag{79b}$$

Then

$$(K^2-\epsilon)\psi(K)=\frac{2Z}{\pi a_0}\int_0^\infty\psi(K')dK'\times\int_0^\infty\frac{zdz}{b^2+z^2}J_0\left(\frac{l'}{K}z\right). \tag{80}$$

The integral over z can be calculated using the integral representation

$$J_0(y)=(2\pi)^{-1}\int_0^{2\pi}e^{iy\cos\varphi}d\varphi.$$

The result is

$$\int_0^\infty\frac{zdz}{b^2+z^2}J_0\left(\frac{l'}{K}z\right)=\tfrac{1}{2}\pi iH_0^{(1)}\left(i\frac{l'b}{K}\right). \tag{81}$$

The function on the right-hand side is also known as the Macdonald function (of real argument). It decreases exponentially for large argument and can generally be written

$$\pi iH_0^{(1)}(ix)=\int_{-\infty}^\infty dy\,e^{-(x^2+y^2)^{\frac{1}{2}}}/(x^2+y^2)^{\frac{1}{2}}=\pi^{-1}\int_{-\infty}^\infty dy\int_{-\infty}^\infty du\,e^{iu}/(x^2+y^2+u^2). \tag{81a}$$

The equation (80) for the determination of ψ becomes now

$$(K^2-\epsilon)\psi(K)=(Z/a_0)\int\psi(K')dK'\times iH_0^{(1)}(il'K^{-1}[(K-K')^2+\alpha^2]^{\frac{1}{2}}). \tag{82}$$

Here we put

$$\psi(K)=a(K)\exp i\int^K\varphi(k)dk. \tag{82a}$$

Neglecting derivatives of a and φ, we obtain

$$\int e^{i\mu x}iH_0^{(1)}\{i(\beta^2+x^2)^{\frac{1}{2}}\}dx=(K^2-\epsilon)a_0(l+\tfrac{1}{2})/ZK \tag{83}$$

with the abbreviations

$$x=(l+\tfrac{1}{2})(K'-K)/K \tag{83a}$$

$$\beta=(l+\tfrac{1}{2})\alpha/K \tag{83b}$$

$$\mu=\varphi(K)K/(l+\tfrac{1}{2}). \tag{83c}$$

Using the integral representations of $H_0^{(1)}$ given in (81a) we find

$$\int_{-\infty}^\infty e^{i\mu x}iH_0^{(1)}\{i(\beta^2+x^2)^{\frac{1}{2}}\}dx=\frac{1}{\pi^2}\int\int\int_{-\infty}^\infty\frac{dxdydu}{\beta^2+x^2+y^2+u^2}e^{i\mu x+iu}. \tag{83d}$$

The integral can be carried out in polar coordinates,

$$x^2+y^2+u^2=r^2,\quad \mu x+u=(\mu^2+1)^{\frac{1}{2}}r\cos\vartheta,$$

and we obtain finally

$$\frac{e^{-(\mu^2+1)^{\frac{1}{2}}\beta}}{(\mu^2+1)^{\frac{1}{2}}}=\frac{(K^2-\epsilon)a_0(l+\frac{1}{2})}{2ZK}\approx(K-\epsilon^{\frac{1}{2}})\frac{a_0(l+\frac{1}{2})}{Z}. \tag{84}$$

[10] Jahnke-Emde, *Funktionentafeln*, 1st edition.

This equation determines μ as a function of K. The upper limit K_2 of the oscillatory region is given by $\mu=0$, i.e.,

$$K_2=\epsilon^{\frac{1}{2}}+\frac{Z}{a_0(l+\frac{1}{2})}e^{-(l+\frac{1}{2})\alpha/K}; \tag{85}$$

the lower limit by $\mu=\infty$, i.e.,

$$K_1=\epsilon^{\frac{1}{2}}. \tag{85a}$$

The oscillatory region is thus narrower the higher l, i.e., the more rapid the oscillation of c as a function of the angle. Its width depends also on the screening constant α (greater α means stronger screening), but this effect becomes less important for higher energy (larger K). As in some examples of §6, the oscillatory region is not symmetrical about the point $K^2=\epsilon$ which corresponds to the unperturbed problem of free electrons. The physical reason for this is that the Coulomb potential is attractive everywhere so that the kinetic energy K^2 is in general larger, but never smaller, than the total energy ϵ. If we define K_2-K_1 as the width of the oscillatory region, it is independent of ϵ, except for the screening. Especially for high Z and small l, the width is very large, namely of the order of the reciprocal radius of the K shell of the atom in question. E.g., for Au ($Z=79$), $l=1$ and $K\gg\alpha$, the width is 100A^{-1}, i.e., about 60 times the elementary spacing in the reciprocal lattice. This shows how much the wave function, even of fast electrons, differs from a plane wave. At the same time, it justifies our replacing all sums over points in the reciprocal lattice by integrals.

It must be pointed out that the considerations of this section actually solve only the problem of an electron in an *atom* with a screened Coulomb field using the wave function in momentum space. The only justification for calling it a solution of the problem of fast electrons in *solids* is that momentum wave functions are specially adapted to solids.

LETTERS TO THE EDITOR

Prompt publication of brief reports of important discoveries in physics may be secured by addressing them to this department. Closing dates for this department are, for the first issue of the month, the eighteenth of the preceding month, for the second issue, the third of the month. Because of the late closing dates for the section no proof can be shown to authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents.

Communications should not in general exceed 600 words in length.

The Nuclear Magnetic Moment of N^{14} *

The $_7N^{14}$ nucleus is one of the few which possess odd atomic number and even atomic weight. Like $_3Li^6$ and $_1H^2$, which are also in this class, its spin is known to be 1. It was suggested by some writers that the moment of $_7N^{14}$ as well as that of $_3Li^6$ should be equal to that of the deuteron, at least in first approximation. These predictions are fulfilled in the case of $_3Li^6$. However, Bacher,[1] as a result of a spectroscopic investigation in which he failed to find any hyperfine structure splitting of lines in the atomic spectrum, set an upper limit of 0.2 for the moment of N^{14}.

Accordingly, the new molecular beam method for the direct measurement of nuclear magnetic moments,[2] previously used[3] for the determination of the moments of Li^6, Li^7 and F^{19} has been applied to a determination of the moment of N^{14}. The method measures the precession frequency, ν, of a nucleus in a uniform magnetic field, H. The moment is obtained from the relation, $\nu=g\mu_0 H/h$ where $g=\mu/i$. The nuclear spin, i, of N^{14} is known[4] to be 1 from the alternating intensities observed in the band spectra of nitrogen.

There was a certain difficulty in finding a compound which contained nitrogen and an alkali atom and which could be vaporized without decomposition. The alkali atom in the molecule is necessary for detection by the method of surface ionization. The molecules finally used were NaCN, KCN and RbCN. Although free alkali atoms were observed in the beam and identified by means of the method of zero moments[5] a sufficient intensity of molecules was obtained for further experimentation. The atoms were not a source of disturbance because at high fields, in the deflecting magnets, they are thrown out of the beam entirely.

Apart from the resonance minimum for the alkali in question a minimum was observed which was common to all of the cyanides. The ratio of ν/H was constant with respect to applied magnetic field over a range from 2100 to 4400 gauss, for NaCN and KCN. The same ν/H minimum was observed in RbCN although in this case the agreement could not be accurately tested. This observed value of ν/H was constant within 1 percent. We attribute this common resonance minimum to N^{14}. It cannot be assigned to C^{12} since its spin is zero. C^{13} must be ruled out because the effect is too large, in view of the low abundance of this isotope. There remains a question whether it can be caused by the rotational moment of the molecule as a whole or to some group such as the CN group common to all the molecules. The former possibility can be ruled out because of the very different moments of inertia of the three molecules making it unlikely that they would all possess the same rotational gyromagnetic ratio. The assumption that the rotation of the CN group, as such (if such a possibility exists at all), has a moment which is the cause of this resonance is ruled out by the fact that ν/H is constant. This means complete decoupling of the hypothetical rotation from the rotation of the molecule as a whole which would be absurd at these low magnetic fields. Moreover, it is very difficult to make a reasonable model of such rotation which would have a Landé g high enough to account for the experimental results.

A further bit of evidence is obtained from the average over-all magnetic moment of the KCN molecule which was found to be about one nuclear magneton. We arrived at this value by observing the deflection of the beam in an inhomogeneous magnetic field. Since these experiments are performed at oven temperatures of about 1000°K the average rotational angular momentum of the molecule or any constituent such as CN is at least $17h/2\pi$. This rotation would result in a magnetic moment of $17\times0.4\sim7$ nuclear magnetons. It would, therefore, be very difficult to assign the resonance to any cause other than a nuclear moment.

The value which we find for $\mu(N^{14})$ is 0.400 ± 0.002 nuclear magneton. This is referred to our published value of 3.265 as the moment of Li^7 which will be our standard, since the ratios are, at present, determined more accurately than the absolute values.

Our value is twice as great as the upper limit assigned by Bacher. However, in view of uncertainties of the calculation of nuclear moments from his data, this fact may be of no further importance. It may be of considerable significance for the nuclear model that the value of $\mu(N^{14})=0.400$ agrees quite well with the results of theoretical considerations of Bethe and Rose[6] and Feenberg and Phillips.[7]

S. MILLMAN
P. KUSCH
I. I. RABI

Department of Physics,
Columbia University,
New York, New York,
November 9, 1938.

* Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

[1] Bacher, Phys. Rev. **43**, 1001 (1933).
[2] Rabi, Zacharias, Millman and Kusch, Phys. Rev. **53**, 318 (1938).
[3] Rabi, Millman, Kusch and Zacharias, Phys. Rev. **53**, 495 (1938).
[4] Ornstein and Van Wijk, Zeits. f. Physik **49**, 315 (1928).
[5] Cohen, Phys. Rev. **46**, 713 (1934).
[6] Bethe and Rose, Phys. Rev. **51**, 205 (1937). See "*Note added in proof.*"
[7] Feenberg and Phillips, Phys. Rev. **51**, 597 (1937).

Evidence of Neutrons in Heavy Particle Showers

Heitler[1] has calculated the cross section for the following reactions which should occur with a frequency that would permit detection.

$$h\nu + P \rightleftarrows Y^+ + N, \quad (1)$$

$$h\nu + N \rightleftarrows Y^- + P. \quad (2)$$

He also states heavy particle showers may be originated by processes similar to the following:

$$Y^+ + N = P + Y^+ + Y^-. \quad (3)$$

In the above, Y is a barytron, N a neutron, and P a proton, while h and ν have their usual significance. While the cross section computed for (1) and (2) only holds for energies of 10^8 ev, we thought it interesting to look for evidence of neutrons in heavy showers. The experimental work done at Denver is as follows: In Fig. 1A, a lead plate, 2 cm in thickness is placed 4 cm above counter 1. Counter 3 is shielded with 10 cm of lead. A block of paraffin alternately occupies positions A and B. With paraffin in position B, ionizing particles ejected from it by neutrons will not trip the counting mechanism; while when the paraffin is in position A, such liberated charged corpuscles may actuate the counting system. The ratio of counts per unit of time with paraffin in position A to those with paraffin in position B is 1.30 ± 0.06, where 0.06 is the probable error of the mean.

The second type of experiment was done with the apparatus shown in Fig. 1B. The three counters in triangular formation were completely surrounded by 10 cm of lead. Readings were taken with 3.4 cm of paraffin in the space above the counters, and with nothing in this space. The ratio of the number of showers per second detected through this system with paraffin to that with no paraffin is 1.27 ∓ 0.07.

In the third experiment the G-M telescope arrangement is illustrated in Fig. 1C. Here the 2 cm block of lead was left above the telescope and respective counts taken with the paraffin in positions A and B. The ratio of counts with paraffin in position A to those when it was between counters 1 and 2 is 1.08 ∓ 0.04.

The arrangement of G-M counters used at Macdonald College is shown in Fig. 2. The shaded areas represent lead and the portions marked P represent paraffin. In all cases, the lead scatterer above the system was of greater than the optimum thickness for shower production, so that the addition of a layer of paraffin to it should decrease the coincidence rate, especially since paraffin contains only very light elements. It was hoped that neutrons formed in the lead scatterer would eject protons from the paraffin to excite the counters. The results obtained up to the present with two sets of counters are given in Table I.

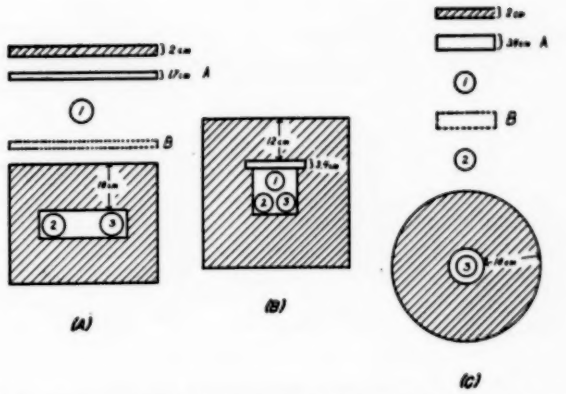

FIG. 1. Arrangements of counters with lead and paraffin absorbers at the University of Denver.

In agreement with Schein and Wilson[2] the ratio (3 : 4) does not differ significantly from unity near sea level. It appears from these results that, (i) the presence of a block of paraffin above a system of counters arranged so as to detect a minimum of one, two or three penetrating ionizing rays increases the counting rate; (ii) this effect is increased by the presence of a lead scatterer above the paraffin; (iii) any increase in counting rate due to the presence of a lead scatterer is magnified by a sheet of paraffin placed immediately below the lead; (iv) paraffin placed immediately above the lower counters in arrangements like b, Fig. 1, also shows an anomalously high absorption for the shower rays, exceeding even the absorption of the penetrating component of the cosmic rays in lead.

We believe that these results support the hypothesis that neutrons or neutrettos are produced in lead by nonionizing radiations. Probably the process is intimately associated with the production of the highly divergent penetrating showers. This would be in agreement with the cloud-chamber observations of Locher.[3]

The high absorption by paraffin may be due to excessive energy losses by the barytrons and protons of the shower in elastic collisions with hydrogen nuclei.

TABLE I. *Dependence of counting rates on the positions of the paraffin and the lead scatterers.*

Counter and absorber arrangement	Counts per hour	Ratio of Rates
1. a, lead at A, with P	36.6 $\mp$0.46	1 : 2, 1.045$\mp$0.017
2. a, lead at B, with P	35.0 $\mp$0.37	3 : 4, 1.018$\mp$0.010
3. a, lead at A, without P	33.6 $\mp$0.24	1 : 3, 1.09 $\mp$0.016
4. a, lead at B, without P	33.0 $\mp$0.23	2 : 4, 1.06 $\mp$0.013
5. b, lead at A, with P	1.17 $\mp$0.06	5 : 6, 1.15 $\mp$0.08
6. b, lead at B, with P counters	1.02 $\mp$0.05	7 : 8, 1.40 $\mp$0.09
7. b, lead at A, without P 1, 2, 3	1.35 $\mp$0.06	5 : 7, 0.87 $\mp$0.06
8. b, lead at B, without P	0.965$\mp$0.05	6 : 8, 1.06 $\mp$0.06
9. b, lead at A, with P	0.176$\mp$0.02	9 : 10, 1.16 $\mp$0.2
10. b, lead at B, with P counters 1, 2, 3, 4	0.151$\mp$0.02	11 : 12, 1.08 $\mp$0.3
11. b, lead at A, without P	0.220$\mp$0.02	9 : 11, 0.80 $\mp$0.12
12. b, lead at B, without P	0.203 $\pm$0.02	10 : 12, 0.75 $\pm$0.12

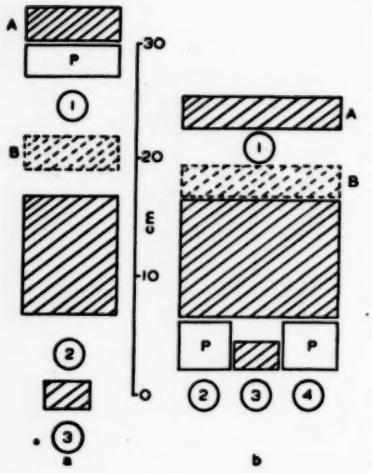

FIG. 2. Arrangements of counters with lead and paraffin absorbers at Macdonald College.

Experiments are being started which will determine how this ratio varies with (i) thickness of paraffin used, (ii) thickness of absorber above lower counter, (iii) thickness of lead scattering block. A more complete study of the apparently high absorption of penetrating cosmic rays by paraffin is also being initiated.

DAROL K. FROMAN
Macdonald College, McGill University,
Quebec, Canada,

J. C. STEARNS
University of Denver,
Denver, Colorado,
November 17, 1938.

[1] W. Heitler, Proc. Roy. Soc. **A927**, 529 (1938).
[2] M. Schein and V. C. Wilson, Phys. Rev. **54**, 304 (1938).
[3] G. L. Locher, Phys. Rev. **44**, 779 (1933); **45**, 296 (1934); **50**, 394 (1936).

Neutron Induced Radioactivity in Columbium

The neutron induced activity in columbium is so weak that it has been reported to be almost completely inactive.[1] According to the survey reported by Pool, Cork and Thornton,[2] two weak periods 7.3 minutes and 3.8 days were found. But neither a test for the sign of the β-rays emitted nor any attempt for the assignment were given yet.

A study of the activity produced in columbium has been made with the Tokyo cyclotron which gives a 3-Mev deuteron beam up to 50 microamperes in intensity.

The results obtained so far are given in Table I.

To test the activities due to some impurities, several runs were made on the special samples which had been purified with extreme care by one of the authors (M. I.) from columbium oxide powder of Kahlbaum.

The relative intensities of 7.5*m*, 66*h* and 11*d* were almost the same as those obtained in other samples. As for the 8*h* period, however, only a trace of its activity was noticed. This fact suggests that this period must be due to a tantalum impurity which is very hard to separate from columbium.

TABLE I. *Results on activity of columbium.*

Slow neutrons (Be+D paraffin)	*W* 7.5±0.5*m*	—	*VW* 66±10*h*	—
Fast neutrons (Li+D)	—	*VW* 8±2*h*	*W* 66±3*h*	*W* 11±1*d*
Sign of β-rays	e^-	e^-	e^-	e^-
Chemical test	—	Cb	Y	Cb
Assignment	Cb^{94}	Ta^{180*}	Y^{90}	Cb^{92}
Reaction	(n, γ)	$(n, 2n)$	(n, α)	$(n, 2n)$
Upper limit derived from K–U plot (Mev)	—	—	2.3±0.3	1±0.2

* Impurity (see reference 1).

The 7.5*m* period is in good agreement with Pool, Cork and Thornton. The 66*h* period is undoubtedly identical with the 3.8*d* period found by Pool and others. A correction made for the 11*d* is responsible for the difference. The K–U plot for this activity is in good agreement with that for Y^{90} produced from Y^{89}.

In conclusion the authors wish to thank Dr. Y. Nishina and Professor K. Kimura for their valuable discussions and encouragement.

One of the authors (R. S.) takes pleasure in expressing his gratitude to Professor E. O. Lawrence for the hospitality with which he was received at the Radiation Laboratory, where many valuable techniques on the cyclotron were obtained and a few runs for preliminary tests on the present problem were made. Thanks are also due to Mr. N. Mori for his assistance. The research has been aided by the grants from the Japan Society for the Promotion of Scientific Research, Oji Paper Manufacturing Company, Japan Wireless Telegraph Company, Mitsui Ho-Onkwai Foundation and Tokyo Electric Light Company.

RYOKICHI SAGANE
SHOJI KOJIMA
GORO MIYAMOTO
MASAO IKAWA
Physics Department (R.S., S.K., G.M.),
Chemistry Department (M.I.),
Tokyo Imperial University,
Tokyo, Japan,
October 22, 1938.

[1] Oldenberg, Phys. Rev. **53**, 35 (1938).
[2] Pool, Cork and Thornton, Phys. Rev. **52**, 239 (1937).

Raman Effect of Dibromofluoromethane

We have observed nine Raman shifts of dibromofluoromethane using our three prism Steinheil spectrograph (dispersion 8A/mm at 4358A). Eight concentric neon-mercury lamps furnished the radiation. No filters were used. The sample was kept at about 40°C by means of an air blast. The results are shown in Table I.

Three types of photographic plates were used: Two Eastman Ortho-Press Plates (exposure 11.5 hours), one Eastman Spectroscopic Plate Type IG (exposure 41.5 hours) and one Eastman Spectroscopic Plate Type IJ (exposure 47 hours). Only on the type IG plate were the Raman lines excited by 5460A noticed and in this case

TABLE I. *Fundamental frequencies of dibromofluoromethane.*

$\Delta\bar{\nu}$ (cm^{-1})	Percent mean deviation	Exciting lines	Number of Readings	Relative Intensity
93.2	0.343	*a b*	4	6
171.4	0.375	*a c*	12	9
295.1	0.251	*a c*	9	5
358.3	0.304	*a b c*	13	9
559.7	0.084	*a c*	7	3
619.4	0.210	*a c*	11	10
700.9	0.034	*c*	3	1
3016.6	0.016	*a c*	5	2
3058.9	0.034	*a c*	4	2

a = 4046A; *b* = 4077A; *c* = 4358A.

they were located too near the plate-edge for accurate measurement and hence they were not included in the average values given in Table I. These lines correspond to 171.4, 295.1 358.3 and 700.9 cm^{-1}. The Raman shifts of 93.2 cm^{-1} scattered by 4358A were noticed on our plates but were partially covered by the halo of the exciting line and hence were too indistinct for measurement.

The sample undergoes photochemical decomposition, turning light brown with the formation of a tarry material, which is insoluble in the ordinary organic solvent (ether, alcohol, acetone, benzene, toluene, carbon tetrachloride, etc.). After the liquid had discolored appreciably, the sample was vacuum-distilled before it was used further. We found that we could expose the sample to the radiation continuously for 47 hours without causing too much background on the photographic plate. The material was furnished by the E. I. du Pont de Nemours Company, and was specified "Laboratory fractionated, boiling range 0.5°C." About 6 cc of the sample were available. We wish to express our gratitude to Dr. A. F. Benning of the Jackson Laboratory for the loan of this substance. Further details concerning these findings will be published later.

GEO. GLOCKLER
J. H. BACHMANN

University of Minnesota,
Minneapolis, Minnesota,
November 18, 1938.

Comments Concerning Anderson's Paper on Contact Difference in Potential Between Barium and Magnesium

In a paper which has just appeared Anderson[1] has described an exhaustive series of experiments on the measurement of the Volta potential between barium and magnesium for the purpose of making a crucial test of the Sommerfeld-Eckart equation, $V_{ab}=\varphi_b-\varphi_a$. For values of φ Anderson used photoelectric work functions determined in this laboratory[2, 3] for magnesium and barium and the value obtained by Mann and DuBridge[4] for magnesium. This choice seems valid since the technique of forming the metal layers was identical in all cases. This technique, developed by the writer as a result of studies on the "clean-up" of gases by alkaline earths, involves a fractional distillation of the metal in high vacuum supplemented by multiple vaporization of the middle fractions. The first fractions "getter" the tube of active gases. This standardized technique for preparing metal layers together with the sensitive electronic method for measuring Volta potentials has enabled Anderson to obtain one of the first conclusive verifications of the theory.

I wish to call attention to certain facts which support Anderson's results and improve the agreement with the Sommerfeld-Eckart equation. First the photoelectric work function for magnesium, 3.60 ev,[2] used by Anderson is an *apparent* work function obtained from the last observable reading of the photoelectric current near the threshold. The true work function must be used in the Sommerfeld-Eckart equation and this can be determined photoelectrically by means of a Fowler plot. Later measurements by the writer of the light intensities used in these experiments permitted the calculation of the true work function by Fowler's method. These results together with a number of others from different phototubes[5] yield the average value 3.67 ± 0.02 ev for the work function of magnesium *deposited on glass*. The "spread" of ±0.02 ev for different tubes is apparently caused by the temperature of the glass when the deposit is made. This value compares well with 3.68 ev obtained by Mann and DuBridge.

Subsequent work[6] on the photoelectric properties of barium has verified the earlier value[3] 2.51 to 2.52 ev used by Anderson. Small changes of the order of 0.02 ev have been observed when barium was deposited on different backing surfaces at different temperatures. With 3.67 and 2.51 ev for the work functions of magnesium and barium respectively, the value for $\varphi_1-\varphi_2$ is 1.16 ev. The unweighted mean for V_{12} obtained by Anderson is 1.15 ev.

The small "spread" of values for the Volta potential which Anderson found for different backing surfaces, which was caused by changes in the work function of magnesium, can be explained, I believe, on the basis of slight changes in the surface crystal structure as Anderson has suggested. This conclusion is based on results[5] from tubes containing in effect two cathodes, one having glass as the backing surface, the other nickel. During the deposition of magnesium the glass was kept at room temperature but the nickel was at a higher temperature, being heated by radiation from the evaporator. The coating on the glass was mirror-like whereas the coating on the nickel resembled a finely etched surface. The work functions were 3.67 and 3.79 ev, respectively. It is significant that Anderson obtained a value of 3.78 ev for the work function of a "macrocrystalline (matte) surface" of magnesium on barium ($\varphi=2.52$ ev).

If the above interpretation is correct, one should, when giving the work function of an evaporated metal, specify the backing surface and its temperature when the deposition of metal occurred unless the actual crystal structure is known.

R. J. CASHMAN

Department of Physics,
Northwestern University,
Evanston, Illinois,
November 12, 1938.

[1] Paul A. Anderson, Phys. Rev. **54**, 753 (1938).
[2] Cashman and Huxford, Phys. Rev. **48**, 734 (1935).
[3] Jamison and Cashman, Phys. Rev. **50**, 624 (1936).
[4] Mann and DuBridge, Phys. Rev. **51**, 120 (1937).
[5] Cashman (unpublished data).
[6] Cashman and Bassoe (in process of publication).

Search for Long or Short-Lived Radio-Chlorine

Of the two radioactive isotopes one expects to form when chlorine is bombarded with deuterons or slow neutrons only one is known. The large capture cross section of chlorine for slow neutrons is not accompanied by a correspondingly intense radioactivity, and it appears probable that a radioactive isotope of chlorine is formed whose half-life is too long or too short to have permitted its detection. A search for a short-lived radioactive body has been carried out by looking for beta-particles in the radiation emitted by chlorine during an irradiation with slow neutrons. Neutrons from a radium-beryllium source passed in succession through 25 cm of lead, 5 cm of paraffin, 5 cm of lead, and a thin layer of C_2Cl_6, which is a conveniently handled compound of chlorine. Close to the C_2Cl_6 a steel thin-walled (0.10 g/cm²) Geiger-Müller counter was placed in such a manner that radiations from the C_2Cl_6 were readily detected, and tin absorbers placed between the C_2Cl_6 and the counter enabled one to distinguish rather sharply between gamma- and beta-rays. Curve 1 (Fig. 1) shows the absorption curve obtained. The slight rise above 100 percent found for small thicknesses of absorber is what one should expect from a thin source of gamma-rays unaccompanied by primary beta-particles. The absence of appreciable numbers of beta-particles indicates the absence of a short-lived beta-emitting radioactive isotope formed by the capture of slow neutrons in chlorine.

For comparison curves 2 and 3 show the results obtained when cadmium and mercury respectively were substituted for the C_2Cl_6. These show the presence of small amounts of some less penetrating component which is probably to be identified with electrons from internally converted gamma-rays. No detectable radioactivity was induced in any of the above substances under the conditions of these experiments. Curve 4 was obtained from a thin foil of silver which had reached radioactive equilibrium with the flux of slow neutrons passing through it. It shows clearly the difference in the type of absorption curve one may obtain when beta-rays as well as gamma-rays are emitted from the irradiated substance. Curve 5 is a conventional absorption curve for the beta+gamma-radiation from the well known 37-min. radio-chlorine (made in the Berkeley cyclotron). It was taken under the same geometrical conditions as the other curves in Fig. 1 and is presented merely to show that even energetic beta-rays (upper limit 6.1 Mev) are distinguishable from gamma-rays of the sort shown in curve 1.

A search for a long-lived radio-chlorine was made as follows. Sodium chloride was bombarded in the Berkeley cyclotron with 100 microampere hours of 5.7 Mev deuterons. An intensive chemical purification of the sample (with great care taken to remove sulfur, phosphorus, iron, lead, potassium and bromine) yielded a product with no detectable radioactivity as measured with an ordinary thin-walled counter nor when measured with an extremely sensitive screen-walled counter of the type described by Libby.[1] If the "missing" radioactive isotope of chlorine is a long-lived beta-emitter, the shortest lifetime consistent with these observations is about 200 years.[2] This calculation has been made by assuming (from analagous reactions

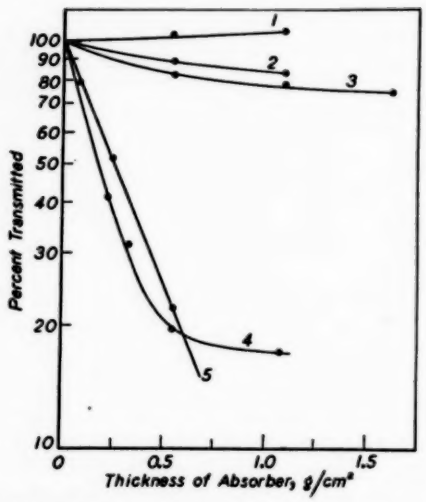

FIG. 1. Absorption of radiation emitted by slow neutron absorbers during irradiation with slow neutrons. Curve 1, C_2Cl_6 (0.14 g/cm²); 2, Cd (0.033 g/cm²); 3, Hg_2O (0.18 g/cm²); 4, Ag (0.037 g/cm²). Curve 5 is an ordinary absorption curve for the radiations from radio-chlorine (37 min. half-life) taken under the same geometrical conditions.

of 5.7 Mev deuterons on neighboring elements) a yield of 1 transmutation per 10^4 deuterons, and by using the first Sargent curve, somewhat extrapolated, to estimate the energy of the beta-rays expected.

It is also possible that Cl^{36} may disintegrate by K-electron capture to form the newly discovered isotope S^{36}.[3] If such is the case, it is possible that no easily detectable radiation would accompany the reaction since the $K\alpha$-radiation from sulfur has a wave-length of about 5.4A and hence lies almost in the so-called vacuum region. It is not necessary that gamma-radiation should be emitted in a K-electron capture since the resultant nucleus may be formed in its ground state. Whether the "missing" isotope of chlorine is very long-lived or indetectable because of K-electron capture is a question which cannot be decided from the present data.

I am greatly indebted to Professor E. O. Lawrence and to members of the staff of the Radiation Laboratory for the preparation of numerous radioactive samples used in these experiments.

D. C. GRAHAME

Department of Chemistry,
University of California,
Berkeley, California,
November 17, 1938.

[1] Libby, Phys. Rev. **46**, 196 (1934).
[2] This surprisingly short minimum life arises from the fact that the beta-radiation assumed has an exceedingly short range in the body of the sample.
[3] Nier, Phys. Rev. **53**, 282 (1938).

The Present Status of the Value of e/m

In 1929, the best value of e/m, from spectroscopic measurements, appeared to be $(1.761\pm0.001)\times10^7$ abs. e.m.u.[1] The best value, from nonspectroscopic (deflection) measurements, was given as 1.769 ± 0.002. The discrepancy was thus nearly three times the sum of the assigned probable errors and appeared so serious that I retained the two separate values, instead of giving one most probable value.

During the years 1930–32 there appeared three new deflection values of e/m, and one new spectroscopic value. All four determinations were mutually consistent and apparently of high accuracy. From them I deduced[2] 1.759 ± 0.001 as the best value of e/m. In other words, the 1929 discrepancy had disappeared, and the error had been shown to lie in the 1929 deflection value.

In the succeeding two years three new values of e/m were obtained, all by chance being just 1.757, and in 1936 I gave[3] 1.75762 ± 0.00026 as the most probable value. It appears now that most of these new "low" values represented preliminary results only, and the final values now available are appreciably higher. In fact Dunnington,[4] in connection with his own beautiful work on e/m, gave 1.7584 ± 0.0003 as the most probable value. He found, however, that a discrepancy of 0.0016 still existed between the weighted averages of the spectroscopic and the deflection measurements, and this, although only one-fifth of the 1929 discrepancy, was still, as a result of the greatly increased accuracy of recent work, almost three times the sum of the assigned probable errors.

At the present time there are available ten precision values of e/m, six spectroscopic by four different methods, and four deflection by three different methods. I find that the discrepancy between the two types of experiment has now shrunk to 0.0006, *just the average deviation to be expected from the assigned probable errors*, and that the final weighted average is 1.75909 ± 0.00024 (external consistency).

To obtain these results I have recalculated each published value (with an occasional slight resulting change) in terms of the following set of auxiliary constants, *viz.*:[5] $c=299776\pm4$ km/sec., $q=0.99993$, $p=1.00048$, and (all on the *physical* scale) $F=9651.31\pm0.80$ abs. e.m.u., $H=1.00813$, $D=2.01473$, $He=4.00389$, $\bar{C}=12.0148$. Each result is weighted according to its probable error, and except as noted, the probable error adopted is just that assigned by the respective investigator. The data are (1 to 6 spectroscopic, 7 to 10 deflection)

(a) Separation of He and H lines
1. $1.7601_5\pm0.0008$[6]

(b) Separation of Hα and Dα lines
2. $1.7581_4\pm0.0004$[7]
3. $1.7579_2\pm0.0004$[8]
4. 1.7592 ± 0.0005[9]

(c) Refraction of x-rays
5. 1.7601 ± 0.0003[10]

(d) Zeeman effect
6. 1.7569 ± 0.0007[11]

(e) Direct velocity measurement
7. 1.7610 ± 0.0010[12]
8. 1.7588 ± 0.0009[13]

(f) Magnetic deflection
9. 1.7597 ± 0.0004[4]

(g) Crossed electric and magnetic fields
10. 1.7571 ± 0.0013[14]

The six spectroscopic results give a weighted average of 1.75895 ± 0.00033 (1.82),[15] the four nonspectroscopic results give 1.75955 ± 0.00033 (0.99), and all ten give 1.75909 ± 0.00024 (1.51) or, considered as the weighted average of the two groups, ±0.00017 (1.07). The nearness to unity of this last ratio, $R_e/R_i=1.07$, shows that the discrepancy between the two groups is just that of the average statistical fluctuation. However, the ratio 1.82 for the six spectroscopic results is unpleasantly large.

That the particular weighting adopted here is relatively unimportant is shown by the fact that the *unweighted* average is 1.75890. As the present most probable value of e/m I recommend $(1.7591\pm0.0003)\times10^7$ abs. e.m.u.

RAYMOND T. BIRGE

University of California,
Berkeley, California,
November 7, 1938.

[1] R. T. Birge, Rev. Mod. Phys. **1**, 1 (1929).
[2] R. T. Birge, Phys. Rev. **42**, 736 (1932).
[3] R. T. Birge, Nature **137**, 187 (1936).
[4] F. G. Dunnington, Phys. Rev. **52**, 475 (1937).
[5] See reference 1 for meaning of symbols.
[6] W. V. Houston, Phys. Rev. **30**, 608 (1927).
[7] C. D. Shane and F. H. Spedding, Phys. Rev. **47**, 33 (1935). The authors' probable error of 0.0003 has been raised to 0.0004, since their result is presumably no more accurate than the other two results by this method.
[8] R. C. Williams, Phys. Rev. **54**, 568 (1938).
[9] W. V. Houston (private communication). This is his present, but not necessarily final result.
[10] J. A. Bearden, Phys. Rev. **54**, 698 (1938).
[11] L. E. Kinsler and W. V. Houston, Phys. Rev. **46**, 533 (1934).
[12] C. T. Perry and E. L. Chaffee, Phys. Rev. **36**, 904 (1930).
[13] F. Kirchner, Ann. d. Physik **12**, 503 (1932).
[14] A. E. Shaw, Phys. Rev. **54**, 193 (1938). The auxiliary constants are not given, and the probable error is that of a least-squares solution, with no additional allowance for other sources of error.
[15] All probable errors are from external consistency, with the ratio of external to internal consistency following in parenthesis. Compare reference 4, page 500.

On the Instability of the Barytron and the Temperature Effect of Cosmic Rays

It is known that the mass absorption of penetrating cosmic rays in air is greater than in earth or water. This effect has been explained by Euler and Heisenberg[1] as due to the instability of the barytrons which form the main part of the penetrating component. These particles are supposed to have a mass M of the order of 150 times the electronic mass and to be of secondary origin. They are produced mainly in the higher levels of the atmosphere by some incident radiation, consisting possibly of electrons.

Following Yukawa, a barytron of energy γMc^2, where $\gamma>>1$, has a mean life $\tau=\gamma\tau_0$, where τ_0 is its mean life when at rest, and is of the order of 10^{-6} sec. In free space, a rapidly moving barytron will travel a mean range $L=c\tau$ before it disintegrates spontaneously into an electron and a neutrino. In dense materials ($\rho\gtrsim1$) the range as defined by the ionization is much less than L, so almost no barytrons decay spontaneously before they come to rest by ionization. But in gases ($\rho\lesssim10^{-3}$) the ionization range is of the order or greater than L, so many barytrons decay before being stopped by ionization, thus producing an apparent additional absorption. Euler and Heisenberg have shown by a detailed analysis that the observed mass absorption anomaly for air and water can be explained by assuming a value of τ_0 of 2.7×10^{-6} sec. The barytrons are supposed to be formed at the maximum of the transition curve, that is, for vertical rays, at a pressure of about 8 cm Hg, and so at a height of about 16 km.

It can easily be seen that the observed decrease of the cosmic-ray intensity with increasing atmospheric tempera-

ture can be explained in a similar way. This decrease results from the greater extension upwards of a warm atmosphere, so that the barytrons are produced at a greater height and so have further to travel to reach sea level, and so have a greater chance of spontaneous decay.

We will simplify the problem by assuming that the barytrons, with which we are concerned, all have an energy[2] of 3×10^9 ev, that is rather less than the measured mean energy of about 4×10^9 ev at sea level in magnetic latitude 54°N. Assuming $M=150\ m_e$, we have $Mc^2=7.5\times10^7$ ev, and therefore $\gamma=40$. With, as before, $\tau_0=2.7\times10^{-6}$ sec., we get $\tau=1.08\times10^{-4}$ sec; and hence $L=32$ km.

If δz is the increase of height of the layer at which the barytrons are formed, due to an increase $\delta\theta$ of mean atmospheric temperature, then the temperature coefficient of the cosmic-ray intensity will be given by

$$\alpha=-\delta z/L\delta\theta.$$

From data given by Humphreys,[3] $\delta z\backsimeq 500m$ for the mean summer-winter temperature difference of 10°C, whence $\alpha=-0.16$ percent per °C. Alternatively instead of using the observed value of $\delta z/\delta\theta$, the atmosphere can be taken as at a uniform temperature $\bar{\theta}$ whence it follows that $\delta z/\delta\theta=z_m/\bar{\theta}$, and so

$$\alpha=-z_m/L\bar{\theta},$$

where z_m is the height of formation of the rays ($z_m=16$ km) and $\bar{\theta}\backsimeq 250$°K. We find $\alpha=0.20$ percent per °C. Both these values are in good agreement with the value of $-0.18+.011$ percent per °C observed by Compton and Turner.[4]

Since near the equator, the incident primary rays are more energetic than in moderate latitudes one would expect the barytrons also to be more energetic, thus having a longer lifetime. So the temperature coefficient at the equator should be lower than at moderate latitude. Owing to the small seasonal variations in the equatorial region, it may be difficult to make the necessary observations to test this. If the temperature coefficient is really lower at the equator, the temperature correction applied by Compton and Turner to obtain the true magnetic latitude effect has been overestimated.

It is probable that the second-order meteorological effects (Corlin[5]) of the cosmic radiation are also explicable on this theory, since Priebsch and Baldauf[6] have pointed out that the main part of these variations can be attributed to a change of density caused by changes of temperature. As an example, it is easy to see why, during some cyclonic depressions, the phase of the rise of cosmic intensity must lag on the pressure drop (Messerschmidt and Pforte[7]). Since the average temperature in front of many depressions is greater than that at the rear, the cosmic-ray intensity will be relatively less, thus producing a phase lag of the intensity maximum compared with the pressure minimum. It should be possible to correlate in some detail the cosmic-ray data with the present-day knowledge of the vertical structure of depressions.

To get a reliable test of this theory of the temperature effect it will be necessary to correlate the observed intensity variations, not as has been done in the past with the local ground temperature, but with the mean temperature of the atmosphere up to a great height.

It does not seem possible to explain the diurnal variation of cosmic-ray intensity in the same way as the temperature effect, since the sign of the effect is opposite to that which one would expect, and further the magnitude does not seem to vary with latitude (Thompson[8]).

P. M. S. BLACKETT

The University,
Manchester, England,
October 10, 1938.

[1] Euler and Heisenberg, Ergeb. d. Exak. Nat. (1938).
[2] Blackett, Proc. Roy. Soc. **159**, 1 (1937).
[3] Humphreys, *Physics of the Air* (Philadelphia, 1930).
[4] Compton and Turner, Phys. Rev. **52**, 799 (1937).
[5] Corlin, Annals of Observatory of Lund, No. 4 (1934).
[6] Priebsch and Baldauf, Ber. Wien. Akad. [II] **145**, 583 (1936).
[7] Messerschmidt and Pforte, Zeits. f. Physik **73**, 677 (1932).
[8] Thompson, Phys. Rev **32**, 140 (1937).